

## PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau

## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6: <b>C12N 15/82, 15/52, 15/62, 15/81, C12P 7/62, A01H 5/00, C08G 63/06</b>	(11) International Publication Number: <b>WO 99/35278</b> (43) International Publication Date: <b>15 July 1999 (15.07.99)</b>	(81) Designated States: <b>AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GR, GM, GW, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SD), OAPI patent (BF, BI, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).</b>  Published <i>With international search report.</i>
(21) International Application Number: <b>PCT/US98/00083</b> (22) International Filing Date: <b>5 January 1998 (05.01.98)</b>  (71) Applicant ( <i>for all designated States except US</i> ): <b>MONSANTO COMPANY [US/US]; 800 North Lindbergh Boulevard, St. Louis, MO 63167 (US).</b>  (72) Inventors; and (75) Inventors/Applicants ( <i>for US only</i> ): <b>POURIER, Yves [CA/CH]; Floréal 3, CH-1006 Lausanne (CH), MITTEN-DORF, Volter [DE/CH]; Faubourg du Lac 25, CH-2000 Neuchâtel (CH).</b>  (74) Agent: <b>KAMMERER, Patricia, A.; Arnold, White &amp; Dutke, P.O. Box 4433, Houston, TX 77210 (US).</b>	(54) Title: <b>BIOSYNTHESIS OF MEDIUM CHAIN LENGTH POLYHYDROXYALKANOATES</b>  (57) Abstract <p>Nucleic acids, proteins, and methods for the biosynthesis of polyhydroxyalkanoate polymer materials are disclosed. In a preferred embodiment, expression of a polyhydroxyalkanoate synthase protein with a peroxisome targeting peptide results in the biosynthesis of medium chain length polyhydroxyalkanoates. In an alternative embodiment, exogenous addition of fatty acids to a plant or cell containing a peroxisome targeted polyhydroxyalkanoate synthase protein leads to the biosynthesis of novel polymeric materials.</p>	

## FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	IS	Iceland	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LJ	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BV	Bolivia	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	St. Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

## BIOSYNTHESIS OF MEDIUM CHAIN LENGTH POLYHYDROXYALKANOATES

### FIELD OF THE INVENTION

The invention relates to the biosynthesis of polymers and more specifically to the biosynthesis of polyhydroxyalkanoate polymers in plants. In particular, a transgenic plant producing peroxisome- or glyoxysome-targeted polyhydroxyalkanoate synthase resulting in the production of polyhydroxyalkanoate materials.

### BACKGROUND OF THE INVENTION

PHAs are bacterial polyesters that accumulate in a wide variety of bacteria. These polymers have properties ranging from stiff and brittle plastics to rubber-like materials, and are biodegradable. Because of these properties, PHAs are an attractive source of nonpolluting plastics and elastomers.

Currently, there are approximately a dozen biodegradable plastics in commercial use that possess properties suitable for producing a number of specialty and commodity products (Lindsay, *Modern Plastics* 2: 62 (1992)). One such biodegradable plastic in the polyhydroxyalkanoate (PHA) family that is commercially important is Biopol™, a random copolymer of 3-hydroxybutyrate (3HB) and 3-hydroxyvalerate (3HV). This bioplastic is used to produce biodegradable molded material (e.g., bottles), films, coatings, and in drug release applications. Biopol™ is produced via a fermentation process employing the bacterium *Alcaligenes eutrophus* (Byrom, *Trends Biotechnol.* 5: 246 (1987)). The current market price is \$6-7/lb, and the annual production is 1,000 tons. By best estimates, this price can be reduced only about 2-fold via fermentation (Poirier et al., *BioTechnology* 13: 142 (1995)). Competitive synthetic plastics such as polypropylene and polyethylene cost about 35-45¢/lb (Layman, *Chem. & Eng. News*, p. 10 (Oct. 31, 1994). The annual global demand for polyethylene alone is about 37 million metric tons (Layman, *Chem. & Eng. News*, p. 10 (Oct. 31, 1994). It is therefore likely that the cost of producing P(3HB-co-3HV) by microbial fermentation will restrict its use to low-volume specialty applications.

Polyhydroxyalkanoate (PHA) is a family of polymers composed primarily of R-3-hydroxyalkanoic acids (Anderson, A. J. & Daves, E. A. *Microbiol. Rev.* 54: 450-472. (1990); Steinbüchel, A. in *Novel Biomaterials from Biological Sources*, ed. Byrom, D. (MacMillan, New York), pp. 123-213. (1991); Poirier, Y., Nawrath, C. & Somerville, C. *BioTechnology* 13: 143-150 (1995)). Polyhydroxybutyrate (PHB) is the most well characterized PHA. High molecular weight PHB is found as intracellular inclusions in a wide variety of bacteria (Steinbüchel, A. in *Novel Biomaterials from Biological Sources*, ed. Byrom, D. (MacMillan, New York), pp. 123-213. (1991)). In *Alcaligenes eutrophus*, PHB typically accumulates to 80% dry weight with inclusions being typically 0.2-1 µm in diameter. Small quantity of PHB oligomers of approximately 150 monomer units are also found associated with membranes of bacteria and eukaryotes, where they form channels permeable to calcium (Reusch, R. N., *Can. J. Microbiol.* 41 (Suppl. 1): 50-54 (1995)). High molecular weight PHAs have the properties of thermoplastics and elastomers. Numerous bacteria and fungi can hydrolyze PHAs to monomers and oligomers, which are metabolized as a carbon source. PHAs have, thus, attracted attention as a potential source of renewable and biodegradable plastics and elastomers. PHB is a highly crystalline polymer with rather poor physical properties, being relatively stiff and brittle (de Koning, G., *Can. J. Microbiol.* 41 (Suppl. 1): 303-309 (1995)). In contrast, PHA copolymers containing monomer units ranging from 3 to 5 carbons for short-chain-length PHA (SCL-PHA), or 6 to 14 carbons for medium-chain-length PHA (MCL-PHA), are less crystalline and more flexible polymers (de Koning, G., *Can. J. Microbiol.* 41 (Suppl. 1): 303-309 (1995)).

PHB has been produced in the plant *Arabidopsis thaliana* expressing the *A. eutrophus* PHB biosynthetic enzymes (Poirier, Y., et al., *Science* 256: 520-523 (1992); Nawrath, C., et al., *Proc. Natl. Acad. Sci. U.S.A.* 91: 12760-12764 (1994)). In plants expressing the PHB pathway in the plastids, leaves accumulated up to 14% PHB per gram dry weight (Nawrath, C., et al., *Proc. Natl. Acad. Sci. U.S.A.* 91: 12760-12764 (1994)). High-level synthesis of PHB in plants opened the possibility of utilizing agricultural crops as a suitable system for the production of PHAs on a large scale and at low cost (Poirier, Y., et al., *BioTechnology* 13: 143-150 (1995); Poirier, Y., et al., *FEMS Microbiol. Rev.* 103: 237-246 (1992); Nawrath, C., et al., *Molecular Breeding* 1: 105-22 (1995)). PHB was also

shown to be synthesized in insect cells expressing a mutant fatty acid synthase (Williams, M. D., et al., *Appl. Environ. Microbiol.* 62: 2540-2546 (1996)), and in yeast expressing the *A. eutrophus* PHB synthase (Leaf, T. A., et al. *Microbiol.* 142: 1169-1180 (1996)).

A number of pseudomonads, including *Pseudomonas putida* and *Pseudomonas aeruginosa*, accumulate MCL-PHAs when cells are grown on alkanolic acids (Anderson, A. J. & Dawes, E. A. *Microbiol. Rev.* 54: 450-472. (1990); Steinbüchel, A. in *Novel Biomaterials from Biological Sources*, ed. Byrom, D. (MacMillan, New York), pp. 123-213. (1991); Poirier, Y. Nawrath, C. & Somerville, C. *BioTechnology* 13: 143-150 (1995)). The nature of the PHA produced is related to the substrate used for growth and is typically composed of monomers which are 2n carbons shorter than the substrate. These studies indicate that MCL-PHAs are synthesized by the PHA synthase from 3-hydroxyacyl-CoA intermediates generated by the  $\beta$ -oxidation of alkanolic acids (Huijberts, G. N. M., et al. *Appl. Environ. Microbiol.* 58: 536-544 (1992); Huijberts, G. N. M., et al., *J. Bacteriol.* 176: 1661-1666 (1994)).

There exists a need for novel methods towards the biosynthesis of polyhydroxyalkanoate materials suitable for commercial applications. Towards this goal, this patent application discloses the materials and methods for the use of a peroxisome targeted polyhydroxyalkanoate synthase protein in the biosynthesis of polyhydroxyalkanoate polymers. Localization in the peroxisomes allow for the utilization of intermediates from the lipid  $\beta$ -oxidation pathway. Plants expressing a *P. aeruginosa* polyhydroxyalkanoate synthase modified for peroxisome targeting produce PHA containing saturated and unsaturated 3-hydroxyalkanoic acids ranging from 6 to 16 carbons. Polyhydroxyalkanoate granules are found within the glyoxysomes or leaf-type peroxisomes of dark- and light-grown plants, respectively, as well as in the vacuoles.

## SUMMARY OF THE INVENTION

The invention is directed towards materials and methods for the biosynthesis of polyhydroxyalkanoate polymers. More particularly, a fusion protein comprising a

polyhydroxyalkanoate synthase protein subunit and a peroxisome targeting protein subunit renders a host cell or plant capable of producing polyhydroxyalkanoate polymer materials.

In one embodiment, the invention provides a non-naturally occurring fusion protein comprising a peroxisome targeting protein subunit and a polyhydroxyalkanoate synthase protein subunit. Generally, the peroxisome targeting protein subunit and the polyhydroxyalkanoate synthase protein subunit may be any subunit suitable for participation in the invention. The peroxisome targeting subunit may be an N-terminal or C-terminal subunit. The N-terminal subunit is preferably PTS2. The C-terminal peroxisome targeting subunit preferably comprises a tripeptide. The first amino acid in the N-terminus to C-terminus direction is preferably S, A, or P. The second amino acid in the N-terminus to C-terminus direction is preferably K, R, S, or H. The third amino acid in the N-terminus to C-terminus direction is L, M, I, or F. More preferably, the C-terminal peroxisome targeting subunit comprises ARM, SRM, SKL, ARL, SRL, PSI, or PRM. The peroxisome targeting subunit is preferably at least 70% identical to SEQ ID NO:14, more preferably at least 80% identical to SEQ ID NO:14, even more preferably at least 90% identical to SEQ ID NO:14, and most preferably is SEQ ID NO:14. The polyhydroxyalkanoate synthase protein subunit is preferably a *Pseudomonas* subunit, and more preferably a *Pseudomonas aeruginosa* subunit. The polyhydroxyalkanoate synthase protein subunit may preferably be either a PHAC1 or PHAC2 subunit. The PHAC1 subunit is preferably at least 70% identical to SEQ ID NO:2, more preferably at least 80% identical to SEQ ID NO:2, even more preferably at least 90% identical to SEQ ID NO:2, and most preferably is SEQ ID NO:2. The PHAC2 subunit is preferably at least 70% identical to SEQ ID NO:4, more preferably at least 80% identical to SEQ ID NO:4, even more preferably at least 90% identical to SEQ ID NO:4, and most preferably is SEQ ID NO:4. The fusion protein is preferably at least 70% identical to SEQ ID NO:18 or SEQ ID NO:20, more preferably at least 80% identical to SEQ ID NO:18 or SEQ ID NO:20, even more preferably at least 90% identical to SEQ ID NO:18 or SEQ ID NO:20, and most preferably is SEQ ID NO:18 or SEQ ID NO:20.

In an alternative embodiment, the invention encompasses a nucleic acid segment encoding a non-naturally occurring fusion protein. The nucleic acid segment preferably comprises a nucleic acid sequence encoding a peroxisome targeting protein subunit, and a

nucleic acid sequence encoding a polyhydroxyalkanoate synthase protein subunit. The nucleic acid sequence encoding a peroxisome targeting protein subunit preferably comprises at least a 6 contiguous nucleic acid sequence from SEQ ID NO:13. The length of the contiguous nucleic acid sequence may be 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, etcetera, 50, 51, 52, etcetera, 100, 101, 102, etcetera, up to and including the entire length of SEQ ID NO:13. The nucleic acid sequence encoding a peroxisome targeting protein subunit is preferably at least 70% identical to SEQ ID NO:13, more preferably at least 80% identical to SEQ ID NO:13, even more preferably at least 90% identical to SEQ ID NO:13, and most preferably is SEQ ID NO:13. The nucleic acid sequence encoding a peroxisome targeting protein subunit preferably hybridizes to SEQ ID NO:13. The nucleic acid sequence encoding a polyhydroxyalkanoate synthase protein subunit preferably comprises at least a 6 contiguous nucleic acid sequence from SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:15, or SEQ ID NO:16. The length of the contiguous nucleic acid sequence may be 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, etcetera, 50, 51, 52, etcetera, 100, 101, 102, etcetera, up to and including the entire length of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:15, or SEQ ID NO:16. The nucleic acid sequence encoding a polyhydroxyalkanoate synthase protein subunit is preferably at least 70% identical to SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:15, or SEQ ID NO:16, more preferably at least 80% identical to SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:15, or SEQ ID NO:16, even more preferably at least 90% identical to SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:15, or SEQ ID NO:16, further preferably is SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:15, or SEQ ID NO:16, and most preferably is SEQ ID NO:15 or SEQ ID NO:16. The nucleic acid sequence encoding a polyhydroxyalkanoate synthase protein subunit preferably hybridizes to SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:15, or SEQ ID NO:16. The encoded peroxisome targeting protein subunit may be an N-terminal or C-terminal peroxisome targeting protein subunit. The encoded N-terminal peroxisome targeting subunit is preferably PTS-2. The encoded C-terminal peroxisome targeting protein subunit preferably comprises a tripeptide. The tripeptide preferably comprises a first amino acid in the N-terminus to C-terminus direction being S, A, or P; a second amino acid in the N-terminus to C-terminus direction being K, R, S, or H; and a third amino acid in the N-terminus to C-terminus direction being L, M, I, or F. The encoded tripeptide preferably is ARM, SRM, SKL, ARL, SRL, PSL, or PRM. The nucleic acid

sequence encoding a polyhydroxyalkanoate synthase protein subunit preferably encodes at least a 5 contiguous amino acid sequence from SEQ ID NO:2 or SEQ ID NO:4. The length of the contiguous nucleic acid sequence may be 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, etcetera, 50, 51, 52, etcetera, 100, 101, 102, etcetera, up to and including the entire length of SEQ ID NO:2 or SEQ ID NO:4. The nucleic acid sequence encoding a polyhydroxyalkanoate synthase protein subunit preferably encodes an amino acid sequence at least 70% identical to SEQ ID NO:2 or SEQ ID NO:4, more preferably at least 80% identical to SEQ ID NO:2 or SEQ ID NO:4, even more preferably at least 90% identical to SEQ ID NO:2 or SEQ ID NO:4, and most preferably is SEQ ID NO:2 or SEQ ID NO:4.

In an alternative embodiment, the invention discloses a recombinant vector comprising in the 5' to 3' direction a) a promoter that directs transcription of a structural nucleic acid sequence encoding a non-naturally occurring fusion protein, wherein the fusion protein comprises a peroxisome targeting protein subunit and a polyhydroxyalkanoate synthase protein subunit, b) a structural nucleic acid sequence encoding a non-naturally occurring fusion protein, wherein the fusion protein comprises a peroxisome targeting protein subunit and a polyhydroxyalkanoate synthase protein subunit, and c) a 3' transcription terminator. The recombinant vector may further comprise a 3' polyadenylation signal sequence that directs the addition of polyadenylate nucleotides to the 3' end of RNA transcribed from the structural nucleic acid coding sequence. The recombinant vector may further comprise a selectable marker. The selectable marker may generally be any selectable marker suitable for the intended host cell or plant, and preferably is a kanamycin resistance marker, a hygromycin resistance marker, or a herbicide resistance marker. The promoter may be constitutive, inducible, tissue specific, or combinations thereof. The constitutive promoter may generally any constitutive promoter suitable for the intended host cell or plant, and preferably is CaMV35S, enhanced CaMV35S, FMV, mas, nos, or ocs. The inducible promoter may generally be any inducible promoter suitable for the intended host cell or plant, and preferably is tac, salicylic acid induced, polyacrylic acid induced, safener induced, heat shock promoter, nitrate induced, hormone induced, or light induced. The tissue specific promoter may generally be any tissue specific promoter suitable for the intended host cell or plant, and preferably is the  $\beta$ -conglycinin 7S promoter,

napin promoter, phascolin promoter, zein promoter, soybean trypsin inhibitor promoter, ACP promoter, stearoyl-ACP desaturase promoter, or oleosin promoter. The nucleic acid sequence encoding a peroxisome targeting protein subunit preferably comprises at least a 6 contiguous nucleic acid sequence from SEQ ID NO:13. The length of the contiguous nucleic acid sequence may be 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, etcetera, 5 50, 51, 52, etcetera, 100, 101, 102, etcetera, up to and including the entire length of SEQ ID NO:13. The nucleic acid sequence encoding a peroxisome targeting protein subunit is preferably at least 70% identical to SEQ ID NO:13, more preferably at least 80% identical to SEQ ID NO:13, even more preferably at least 90% identical to SEQ ID NO:13, and most 10 preferably is SEQ ID NO:13. The nucleic acid sequence encoding a peroxisome targeting protein subunit preferably hybridizes to SEQ ID NO:13. The encoded peroxisome targeting protein subunit may be an N-terminal or C-terminal peroxisome targeting protein subunit. The encoded N-terminal peroxisome targeting subunit is preferably PTS-2. The encoded C-terminal peroxisome targeting protein subunit preferably comprises a tripeptide. The 15 tripeptide preferably comprises a first amino acid in the N-terminus to C-terminus direction being S, A, or P; a second amino acid in the N-terminus to C-terminus direction being K, R, S, or H; and a third amino acid in the N-terminus to C-terminus direction being L, M, I, or F. The encoded tripeptide preferably is ARM, SRM, SKL, ARL, SRL, PSI, or PRM. The encoded polyhydroxyalkanoate synthase protein subunit is preferably a *Pseudomonas* subunit, and more preferably is a *Pseudomonas aeruginosa* subunit. The nucleic acid sequence encoding a polyhydroxyalkanoate synthase protein subunit preferably comprises at least a 6 contiguous nucleic acid sequence from SEQ ID NO:1, SEQ ID NO:3, SEQ ID 20 NO:15, or SEQ ID NO:16. The length of the contiguous nucleic acid sequence may be 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, etcetera, 50, 51, 52, etcetera, 100, 101, 102, etcetera, up to and including the entire length of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:15, or SEQ ID NO:16. The nucleic acid sequence encoding a polyhydroxyalkanoate synthase protein subunit is preferably at least 70% identical to SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:15, or SEQ ID NO:16, more preferably at least 80% identical to SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:15, or SEQ ID NO:16, even more preferably at least 90% 25 identical to SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:15, or SEQ ID NO:16, further preferably is SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:15, or SEQ ID NO:16, and most

preferably is SEQ ID NO:15 or SEQ ID NO:16. The nucleic acid sequence encoding a polyhydroxyalkanoate synthase protein subunit preferably hybridizes to SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:15, or SEQ ID NO:16. The nucleic acid sequence encoding a polyhydroxyalkanoate synthase protein subunit preferably encodes at least a 5 contiguous amino acid sequence from SEQ ID NO:2 or SEQ ID NO:4. The length of the contiguous nucleic acid sequence may be 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, etcetera, 5 50, 51, 52, etcetera, 100, 101, 102, etcetera, up to and including the entire length of SEQ ID NO:2 or SEQ ID NO:4. The nucleic acid sequence encoding a polyhydroxyalkanoate synthase protein subunit preferably encodes an amino acid sequence at least 70% identical to SEQ ID NO:2 or SEQ ID NO:4, more preferably at least 80% identical to SEQ ID NO:2 or SEQ ID NO:4, even more preferably at least 90% identical to SEQ ID NO:2 or SEQ ID NO:4, and most preferably is SEQ ID NO:2 or SEQ ID NO:4. The structural nucleic acid sequence preferably comprises SEQ ID NO:17 or SEQ ID NO:19, and preferably encodes 10 SEQ ID NO:18 or SEQ ID NO:20.

15 In an alternative embodiment, the invention encompasses a recombinant host cell comprising a nucleic acid segment encoding a non-naturally occurring fusion protein, wherein the nucleic acid segment comprises a nucleic acid sequence encoding a peroxisome targeting protein subunit and a nucleic acid sequence encoding a polyhydroxyalkanoate synthase protein subunit. The recombinant host cell may generally be any type of host cell, and preferably is a fungal or plant host cell. The fungal cell is generally any type of fungal cell, and preferably a *Schizosaccharomyces pombe*, *Streptomyces rimofaciens*, *Fusarium*, *Aspergillus niger*, or *Saccharomyces cerevisiae* cell. The plant cell is generally any type of plant cell, and preferably an alfalfa, banana, barley, bean, cabbage, canola/oilseed rape, carrot, castorbean, celery, clover, coconut, corn, cotton, cucumber, linseed, melon, olive, 25 palm, parsnip, pea, peanut, pepper, potato, radish, rapeseed, rice, soybean, spinach, sunflower, tobacco, tomato, or wheat cell. The recombinant host cell may further comprise a nucleic acid segment encoding an acyl-ACP thioesterase, a fatty acyl hydroxylase, a yeast multifunctional protein (MFP), or an hydroxyacyl-CoA epimerase.

A further alternative embodiment describes a genetically transformed plant cell comprising in the 5' to 3' direction: a) a promoter to direct transcription of a structural nucleic acid sequence encoding a non-naturally occurring fusion protein, wherein the structural nucleic acid sequence comprises: i) a nucleic acid sequence encoding a peroxisome targeting protein subunit, and ii) a nucleic acid sequence encoding a polyhydroxyalkanoate synthase protein subunit, b) a structural nucleic acid sequence encoding a non-naturally occurring fusion protein, wherein the structural nucleic acid sequence comprises: i) a nucleic acid sequence encoding a peroxisome targeting protein subunit, and ii) a nucleic acid sequence encoding a polyhydroxyalkanoate synthase protein subunit, c) a 3' transcription terminator sequence, and d) a 3' polyadenylation signal sequence that directs the addition of polyadenylate nucleotides to the 3' end of RNA transcribed from the structural nucleic acid coding sequence. The plant cell is generally any type of plant cell, and preferably an alfalfa, banana, barley, bean, cabbage, canola/oi-seed rape, carrot, castorbean, celery, clover, coconut, corn, cotton, cucumber, linseed, melon, olive, palm, parsnip, pea, peanut, pepper, potato, radish, rapeseed, rice, soybean, spinach, sunflower, tobacco, tomato, or wheat cell. The plant cell may further comprise a nucleic acid segment encoding an acyl-ACP thioesterase, a fatty acyl hydroxylase, a yeast multifunctional protein (MFP), or an hydroxyacyl-CoA epimerase.

An additional embodiment describes a genetically transformed plant comprising in the 5' to 3' direction: a) a promoter to direct transcription of a structural nucleic acid sequence encoding a non-naturally occurring fusion protein, wherein the structural nucleic acid sequence comprises: i) a nucleic acid sequence encoding a peroxisome targeting protein subunit, and ii) a nucleic acid sequence encoding a polyhydroxyalkanoate synthase protein subunit, b) a structural nucleic acid sequence encoding a non-naturally occurring fusion protein, wherein the structural nucleic acid sequence comprises: i) a nucleic acid sequence encoding a peroxisome targeting protein subunit, and ii) a nucleic acid sequence encoding a polyhydroxyalkanoate synthase protein subunit, c) a 3' transcription terminator sequence, and d) a 3' polyadenylation signal sequence that directs the addition of polyadenylate nucleotides to the 3' end of RNA transcribed from the structural nucleic acid coding sequence. The plant may generally be any type of plant, and preferably an alfalfa, banana,

barley, bean, cabbage, canola/oi-seed rape, carrot, castorbean, celery, clover, coconut, corn, cotton, cucumber, linseed, melon, olive, palm, parsnip, pea, peanut, pepper, potato, radish, rapeseed, rice, soybean, spinach, sunflower, tobacco, tomato, or wheat plant. The promoter may be constitutive, inducible, tissue specific, or combinations thereof. The constitutive promoter may generally any constitutive promoter suitable for the intended plant, and preferably is CaMV35S, enhanced CaMV35S, FMV, mas, nos, or ocs. The inducible promoter may generally be any inducible promoter suitable for the intended plant, and preferably is tac, salicylic acid induced, polyacrylic acid induced, safener induced, heat shock promoter, nitrate induced, hormone induced, or light induced. The tissue specific promoter is generally any tissue specific promoter, and preferably is the  $\beta$ -conglycinin 7S promoter, napin promoter, phaseolin promoter, zein promoter, soybean trypsin inhibitor promoter, ACP promoter, stearyl-ACP desaturase promoter, or oleosin promoter. The plant may further comprise a nucleic acid segment encoding an acyl-ACP thioesterase, a fatty acyl hydroxylase, a yeast multifunctional protein (MFP), or an hydroxyacyl-CoA epimerase.

The invention describes a method for preparing host cells useful to produce a non-naturally occurring fusion protein comprising the steps of: a) selecting a host cell b) transforming the selected host cell with a recombinant vector having a structural nucleic acid sequence encoding a non-naturally occurring fusion protein, wherein the structural nucleic acid sequence comprises: i) a nucleic acid sequence encoding a peroxisome targeting protein subunit, and ii) a nucleic acid sequence encoding a polyhydroxyalkanoate synthase protein subunit, and c) obtaining transformed host cells. The vector may further comprise a selectable marker. The selectable marker may generally be any selectable marker suitable for use in the intended host cell, and more preferably for plants is a kanamycin resistance marker, a hygromycin resistance marker, or a herbicide resistance marker. The host cell may generally be any type of cell, and preferably is a fungal or plant cell. The fungal cell may generally be any type of fungal cell, and more preferably is a *Schizosaccharomyces pombe*, *Streptomyces rimofaciens*, *Fusarium*, *Aspergillus niger*, or *Saccharomyces cerevisiae* cell. The plant cell may generally be any type of plant cell, and more preferably is an alfalfa, banana, barley, bean, cabbage, canola/oi-seed rape, carrot,

castorbean, celery, clover, coconut, corn, cotton, cucumber, linseed, melon, olive, palm, parsnip, pea, peanut, pepper, potato, tomato, radish, rapeseed, rice, soybean, spinach, sunflower, tobacco, tomato, or wheat cell.

The invention further describes a method of preparing a transformed plant useful to produce a non-naturally occurring fusion protein comprising the steps of: a) selecting a host plant cell b) transforming the selected host cell with a recombinant vector having a structural nucleic acid sequence encoding a non-naturally occurring fusion protein, wherein the structural nucleic acid sequence comprises: i) a nucleic acid sequence encoding a peroxisome targeting protein subunit; and ii) a nucleic acid sequence encoding a polyhydroxyalkanoate synthase protein subunit; c) obtaining transformed host plant cells; and d) regenerating the transformed host plant cells. The vector may further comprise a selectable marker. The selectable marker may generally be any selectable marker suitable for use in the intended host cell, and more preferably is a kanamycin resistance marker, a hygromycin resistance marker, or a herbicide resistance marker. The host plant cell may generally be any type of plant cell, and more preferably is an alfalfa, banana, barley, bean, cabbage, canola/oilseed rape, carrot, castorbean, celery, clover, coconut, corn, cotton, cucumber, linseed, melon, olive, palm, parsnip, pea, peanut, pepper, potato, tomato, radish, rapeseed, rice, soybean, spinach, sunflower, tobacco, tomato, or wheat cell. The invention also encompasses the plant made by the above described methods.

A preferred embodiment is a method for the preparation of a polyhydroxyalkanoate, comprising the steps of: a) obtaining a cell capable of producing a non-naturally occurring fusion protein, wherein the fusion protein comprises: i) a peroxisome targeting protein subunit; and ii) a polyhydroxyalkanoate synthase protein subunit; b) establishing a culture of the cell; and c) culturing the cell under conditions suitable for the production of the polyester. The method may further comprise isolating the polyhydroxyalkanoate from the cultured cell. The culture may further comprise fatty acids, and more preferably natural fatty acids, non-natural or synthetic fatty acids, or mixtures thereof. The cell may generally be any type of cell, and preferably is a fungal or plant cell. The fungal cell may generally be any type of fungal cell, and more preferably is a *Schizosaccharomyces pombe*, *Sreptomyces*

*rimofaciens*, *Fusarium*, *Aspergillus niger*, or *Saccharomyces cerevisiae* cell. The plant cell may generally be any type of plant cell, and more preferably is an alfalfa, banana, barley, bean, cabbage, canola/oilseed rape, carrot, castorbean, celery, clover, coconut, corn, cotton, cucumber, linseed, melon, olive, palm, parsnip, pea, peanut, pepper, potato, tomato, radish, rapeseed, rice, soybean, spinach, sunflower, tobacco, tomato, or wheat cell. The polyhydroxyalkanoate isolated from the cell may generally be any type of polyhydroxyalkanoate, and preferably comprises 3-hydroxyhexanoic acid (H:6), 3-hydroxyoctanoic acid (H:8), 3-hydroxydecanoic acid (H:10), 3-hydroxydodecanoic acid (H:12), 3-hydroxytetradecanoic acid (H:14), 3-hydroxyhexadecanoic acid (H:16), 3-hydroxyheptanoic acid (H:7), 3-hydroxynonanoic acid (H:9), 3-hydroxyundecanoic acid (H:11), 3-hydroxytridecanoic acid (H:13), 3-hydroxyhexadecatrienoic acid (H16:3), 3-hydroxyhexadecadienoic acid (H16:2), 3-hydroxyhexadecenoic acid (H16:1), 3-hydroxytetradecatrienoic acid (H14:3), 3-hydroxytetradecadienoic acid (H14:2), 3-hydroxytetradecenoic acid (H14:1), 3-hydroxydodecadienoic acid (H12:2), 3-hydroxydodecenoic acid (H12:1), 3-hydroxyoctenoic acid (H8:1), 4-hydroxydecanoic acid, 8-methyl-3-hydroxynonanoic acid, or 6-methyl-3-hydroxyheptanoic acid monomers.

In a further preferred embodiment, the invention presents a method for the preparation of a polyhydroxyalkanoate, comprising the steps of: a) obtaining a plant capable of producing a non-naturally occurring fusion protein, wherein the fusion protein comprises: i) a peroxisome targeting protein subunit; and ii) a polyhydroxyalkanoate synthase protein subunit; and c) growing the plant under conditions suitable for the production of the polyhydroxyalkanoate. The method may further comprise the step of isolating the polyhydroxyalkanoate from the plant. The method may further comprise supplementing the plant with natural fatty acids, non-natural fatty acids, or mixtures thereof. The plant may generally be any type of plant, and preferably is an alfalfa, banana, barley, bean, cabbage, canola/oilseed rape, carrot, castorbean, celery, clover, coconut, corn, cotton, cucumber, linseed, melon, olive, palm, parsnip, pea, peanut, pepper, potato, tomato, radish, rapeseed, rice, soybean, spinach, sunflower, tobacco, tomato, or wheat plant. The polyhydroxyalkanoate isolated from the plant may generally be any type of polyhydroxyalkanoate, and preferably comprises 3-hydroxyhexanoic acid (H:6), 3-

hydroxyoctanoic acid (H:8), 3-hydroxydecanoic acid (H:10), 3-hydroxydodecanoic acid (H:12), 3-hydroxytetradecanoic acid (H:14), 3-hydroxyhexadecanoic acid (H:16), 3-hydroxyheptanoic acid (H:7), 3-hydroxynonanoic acid (H:9), 3-hydroxyundecanoic acid (H:11), 3-hydroxytridecanoic acid (H:13), 3-hydroxyhexadecatrienoic acid (H16:3), 3-hydroxyhexadecadienoic acid (H16:2), 3-hydroxyhexadecenoic acid (H16:1), 3-hydroxytetradecatrienoic acid (H14:3), 3-hydroxytetradecadienoic acid (H14:2), 3-hydroxytetradecenoic acid (H14:1), 3-hydroxydodecadatrienoic acid (H12:2), 3-hydroxydodecenoic acid (H12:1), 3-hydroxyoctenoic acid (H8:1), 4-hydroxydecanoic acid, 8-methyl-3-hydroxynonanoic acid, or 6-methyl-3-hydroxyheptanoic acid monomers.

The invention further encompasses plants containing polyhydroxyalkanoates, wherein the polyhydroxyalkanoate comprises 3-hydroxyhexanoic acid (H:6), 3-hydroxyoctanoic acid (H:8), 3-hydroxydecanoic acid (H:10), 3-hydroxydodecanoic acid (H:12), 3-hydroxytetradecanoic acid (H:14), 3-hydroxyhexadecanoic acid (H:16), 3-hydroxyheptanoic acid (H:7), 3-hydroxynonanoic acid (H:9), 3-hydroxyundecanoic acid (H:11), 3-hydroxytridecanoic acid (H:13), 3-hydroxyhexadecatrienoic acid (H16:3), 3-hydroxyhexadecadienoic acid (H16:2), 3-hydroxyhexadecenoic acid (H16:1), 3-hydroxytetradecatrienoic acid (H14:3), 3-hydroxytetradecadienoic acid (H14:2), 3-hydroxytetradecenoic acid (H14:1), 3-hydroxydodecadatrienoic acid (H12:2), 3-hydroxydodecenoic acid (H12:1), 3-hydroxyoctenoic acid (H8:1), 4-hydroxydecanoic acid, 8-methyl-3-hydroxynonanoic acid, or 6-methyl-3-hydroxyheptanoic acid monomers.

In an alternative embodiment, the invention describes polyhydroxyalkanoates comprising 3-hydroxyhexadecatrienoic acid (H16:3), 3-hydroxyhexadecadienoic acid (H16:2), 3-hydroxytetradecatrienoic acid (H14:3), or 3-hydroxydodecadatrienoic acid (H12:2) monomers.

## DESCRIPTION OF THE FIGURES

The following figure forms part of the present specification and is included to further demonstrate certain aspects of the present invention. The invention may be better

-13-

understood by reference to the figure in combination with the detailed description of specific embodiments presented herein.

Figure 1: GC-MS analysis of PHA in transgenic plants. Trans-esterified chloroform extracts from phaC1-transformed line 3.3 (1A, 1B) and vector-transformed line 21 (1C, 1D) were analyzed. In panels 1A and 1C, the total ion chromatogram is presented, while on panel 1B and 1D, only ions with a mass-to-charge ratio of 103 are shown.

## DESCRIPTION OF THE SEQUENCE LISTINGS

The following sequence listings form part of the present specification and are included to further demonstrate certain aspects of the present invention. The invention may be better understood by reference to one or more of these sequence listings in combination with the detailed description of specific embodiments presented herein.

SEQ ID NO:1	Wild type PHA synthase C1 nucleic acid sequence.
SEQ ID NO:2	Wild type PHA synthase C1 protein sequence.
SEQ ID NO:3	Wild type PHA synthase C2 nucleic acid sequence.
SEQ ID NO:4	Wild type PHA synthase C2 protein sequence.
SEQ ID NO:5	Forward PCR primer for PHA synthase C1 fusion sequence.
SEQ ID NO:6	Reverse PCR primer for PHA synthase C1 fusion sequence.
SEQ ID NO:7	Forward PCR primer for PHA synthase C2 fusion sequence.
SEQ ID NO:8	Reverse PCR primer for PHA synthase C2 fusion sequence.
SEQ ID NO:9	Wild type isocitrate lyase nucleic acid sequence.
SEQ ID NO:10	Wild type isocitrate lyase protein sequence.
SEQ ID NO:11	Forward PCR primer for isocitrate lyase fusion sequence.
SEQ ID NO:12	Reverse PCR primer for isocitrate lyase fusion sequence.
SEQ ID NO:13	Nucleic acid sequence encoding the isocitrate lyase petoxisome targeting protein subunit.
SEQ ID NO:14	Isocitrate lyase petoxisome targeting protein subunit.
SEQ ID NO:15	PHA synthase C1 nucleic acid sequence with plant preferred codon.

-14-



SEQ ID NO:16 PHA synthase C2 nucleic acid sequence with plant preferred codon.

SEQ ID NO:17 Nucleic acid sequence encoding PHA synthase C1 and isocitrate lyase fusion protein.

SEQ ID NO:18 PHA synthase C1 and isocitrate lyase fusion protein.

SEQ ID NO:19 Nucleic acid sequence encoding PHA synthase C2 and isocitrate lyase fusion protein.

SEQ ID NO:20 PHA synthase C2 and isocitrate lyase fusion protein.

SEQ ID NO:21 PCR amplified nucleic acid sequence encoding wild type *Candida albicans* MFP.

SEQ ID NO:22 Wild type *Candida albicans* MFP protein.

SEQ ID NO:23 PCR amplified nucleic acid sequence encoding SKL mutant *Candida albicans* MFP.

SEQ ID NO:24 *Candida albicans* MFP protein with SKL substitution for AKI.

SEQ ID NO:25 PCR amplified nucleic acid sequence encoding mutant *Candida albicans* MFP lacking AKI sequence.

SEQ ID NO:26 *Candida albicans* MFP protein lacking AKI sequence.

# DEFINITIONS

The following definitions are provided in order to aid those skilled in the art in understanding the detailed description of the present invention.

"AcyI-ACP thioesterase" refers to proteins which catalyze the hydrolysis of acyl-ACP thioesters.

"C-terminal region" refers to the region of a peptide, polypeptide, or protein chain from the middle thereof to the end that carries the amino acid having a free a carboxyl group (the C-terminus).

"CoA" refers to coenzyme A.

The phrases "coding sequence", "open reading frame", and "structural sequence" refer to the region of continuous sequential nucleic acid triplets encoding a protein, polypeptide, or peptide sequence.

The term "encoding DNA" or "encoding nucleic acid" refers to chromosomal nucleic acid, plasmid nucleic acid, cDNA, or synthetic nucleic acid which codes on expression for any of the proteins or fusion proteins discussed herein.

"Fatty acyl hydroxylase" refers to proteins which catalyze the conversion of fatty acids to hydroxylated fatty acids.

The term "gene" refers to chromosomal DNA, plasmid DNA, cDNA, synthetic DNA, or other DNA that encodes a peptide, polypeptide, protein, or RNA molecule, and regions flanking the coding sequence involved in the regulation of expression.

The term "genome" as it applies to bacteria encompasses both the chromosome and plasmids within a bacterial host cell. Encoding DNAs of the present invention introduced into bacterial host cells can therefore be either chromosomally-integrated or plasmid-localized. The term "genome" as it applies to plant cells encompasses not only chromosomal DNA found within the nucleus, but organelle DNA found within subcellular components of the cell. DNAs of the present invention introduced into plant cells can therefore be either chromosomally-integrated or organelle-localized.

"Glyoxysome" and "peroxisome" refer to the same organelle in a plant. Glyoxysome refers to a type of peroxisome found in germinating seedlings, senescing tissues, or in dark-grown tissues. Glyoxysomes and peroxisomes contain enzymes responsible for the conversion of lipids to carbohydrates.

"Identity" refers to the degree of similarity between two nucleic acid or protein sequences. An alignment of the two sequences is performed by a suitable computer

program. A widely used and accepted computer program for performing sequence alignments is CLUSTALW v1.6 (Thompson, et al. *Nucl. Acids Res.*, 22: 4673-4680 (1994)). The number of matching bases or amino acids is divided by the total number of bases or amino acids, and multiplied by 100 to obtain a percent identity. For example, if two 580 base pair sequences had 145 matched bases, they would be 25 percent identical. If the two compared sequences are of different lengths, the number of matches is divided by the shorter of the two lengths. For example, if there were 100 matched amino acids between 200 and a 400 amino acid proteins, they are 50 percent identical with respect to the shorter sequence.

The terms "microbe" or "microorganism" refer to algae, bacteria, fungi, and protozoa.

"N-terminal region" refers to the region of a peptide, polypeptide, or protein chain from the amino acid having a free amino group to the middle of the chain.

"Nucleic acid" refers to ribonucleic acid (RNA) and deoxyribonucleic acid (DNA).

A "nucleic acid segment" is a nucleic acid molecule that has been isolated free of total genomic DNA of a particular species, or that has been synthesized. Included with the term "nucleic acid segment" are DNA segments, recombinant vectors, plasmids, cosmids, phagemids, phage, viruses, etcetera.

"Overexpression" refers to the expression of a polypeptide or protein encoded by a DNA introduced into a host cell, wherein said polypeptide or protein is either not normally present in the host cell, or wherein said polypeptide or protein is present in said host cell at a higher level than that normally expressed from the endogenous gene encoding said polypeptide or protein.

The term "plastid" refers to the class of plant cell organelles that includes amyloplasts, chloroplasts, chromoplasts, elaioplasts, coplasts, etioplasts, leucoplasts, and proplastids. These organelles are self-replicating, and contain what is commonly referred to

as the "chloroplast genome," a circular DNA molecule that ranges in size from about 120 to about 217 kb, depending upon the plant species, and which usually contains an inverted repeat region (Fosket, Plant Growth and Development, Academic Press, Inc., San Diego, CA, p. 132 (1994)).

"Polyadenylation signal" or "polyA signal" refers to a nucleic acid sequence located 3' to a coding region that directs the addition of adenylate nucleotides to the 3' end of the mRNA transcribed from the coding region.

The term "polyhydroxyalkanoate (or PHA) synthase" refers to enzymes that convert hydroxyacyl-CoAs to polyhydroxyalkanoates and free CoA.

The term "promoter" or "promoter region" refers to a nucleic acid sequence, usually found upstream (5') to a coding sequence, that controls expression of the coding sequence by controlling production of messenger RNA (mRNA) by providing the recognition site for RNA polymerase and/or other factors necessary for start of transcription at the correct site. As contemplated herein, a promoter or promoter region includes variations of promoters derived by means of ligation to various regulatory sequences, random or controlled mutagenesis, and addition or duplication of enhancer sequences. The promoter region disclosed herein, and biologically functional equivalents thereof, are responsible for driving the transcription of coding sequences under their control when introduced into a host as part of a suitable recombinant vector, as demonstrated by its ability to produce mRNA.

"Protein subunit" refers to a protein sequence that is part of a fusion protein. Examples are  $\beta$ -galactosidase, FLAG, green fluorescent protein, and in the instant invention, polyhydroxyalkanoate synthase, and a peroxisome or glyoxysome targeting peptide.

"PTS2" refers to an N-terminal protein subunit having the sequence (RKX)(L/Q/I)XXXXXXXX(H/Q)L, wherein X is any amino acid.

"Regeneration" refers to the process of growing a plant from a plant cell (e.g., plant protoplast or explant).

"Transformation" refers to a process of introducing an exogenous nucleic acid sequence (e.g., a vector, recombinant nucleic acid molecule) into a cell or protoplast in which that exogenous nucleic acid is incorporated into a chromosome or is capable of autonomous replication.

A "transformed cell" or "transgenic cell" is a cell whose DNA has been altered by the introduction of an exogenous nucleic acid molecule into that cell.

A "transformed plant" or "transgenic plant" is a plant whose DNA has been altered by the introduction of an exogenous nucleic acid molecule into that plant, or by the introduction of an exogenous nucleic acid molecule into a plant cell from which the plant was regenerated or derived.

The following examples are included to demonstrate preferred embodiments of the invention. It should be appreciated by those of skill in the art that the techniques disclosed in the examples which follow represent techniques discovered by the inventors to function well in the practice of the invention, and thus can be considered to constitute preferred modes for its practice. However, those of skill in the art should, in light of the present disclosure, appreciate that many changes can be made in the specific embodiments which are disclosed and still obtain a like or similar result without departing from the spirit and scope of the invention.

## EXAMPLES

### EXAMPLE 1: Plant material

*Arabidopsis thaliana*, race Columbia, was transformed by the vacuum infiltration method (Bechtold, N., et al., *C.R. Acad. Sci. Paris* 316: 1194-1199 (1993)). Transformants

were selected on media containing Murashige and Skoog salts ("MS", Murashige, T. and Skoog, F., *Physiol. Plant.* 15: 473-497 (1962)), 1% (w/v) sucrose, 0.7% (w/v) agar and 50 µg/mL kanamycin. Kanamycin-resistant plants were subsequently transferred to soil and grown under continuous fluorescent light at 19°C. In some experiments, plants were grown under constant agitation (100 rpm) for 1-2 weeks in liquid media containing MS salts and 2% sucrose.

### EXAMPLE 2: Cloning of peroxisomally targeted PHA synthases C1 and C2

The *phaC1* and *phaC2* genes were obtained from Steinbüchel (Timm, A. and Steinbüchel, A., *Eur. J. Biochem.* 209: 14-30 (1992), GenBank Accession Number X66592). PCR was used to amplify the genes and to modify their 5'- and 3'-termini as follows: At the 5'-end the codons encoding the serine-2 and the arginine-2 residue of *phaC1* and *phaC2*, respectively, were modified to conform more closely with the general codon preferences of *A. thaliana* (Meyerowitz, E. M. in *Methods in Arabidopsis research*, eds. Konec, C., Chua, N.-H. & Schell, J. (World Scientific Publishing, Singapore), pp. 100-119 (1992)). At the 3'-end the sequences were modified to obtain suitable cloning sites and to delete the stop codons to enable the construction of chimerical fusions with the peroxisomal targeting sequence.

The carboxy-terminal 35 amino acid residues of the isocitrate lyase gene (ICL) (Olsen, L.J., et al., *Plant Cell* 5: 941-952 (1993), GenBank Accession Number Y13356) from *Brassica napus* were used as targeting sequence for the PHA synthases C1 and C2. It has been shown previously that this sequence was sufficient to ensure the peroxisomal localization of the chloramphenicol acetyl transferase (CAT) to the peroxisomes in *A. thaliana* (Comai, L. et al., *The Plant Cell* 1: 293-300 (1989); Olsen, L. J. et al., *The Plant Cell* 5: 941-952 (1993); Zhang, J. Z. et al., *Mol. Gen. Genet.* 238: 177-184 (1993)). A PCR product encoding the ICL targeting sequence was cloned into the vector pART7 (Gleaves, A.P., *Plant Mol. Biol.* 20: 1202-1207 (1992), GenBank Accession Number X69707). The PCR products containing the *phaC1* or *phaC2* genes were cloned 5'-upstream of the ICL sequence to produce a contiguous open reading frame encoding the targeted fusion proteins.

The 5' and 3'-ends of the genes in the resulting plasmids pART7\_phaC1\_ICL and pART7\_phaC2\_ICL were sequenced to verify the modifications.

The PHA accumulation-deficient mutant *Pseudomonas putida* KT2440 NK2:3 was obtained from Steinbüchel for complementation studies to verify the enzyme activities of the modified PHA synthases C1 and C2. The *phaC1\_ICL* and *phaC2\_ICL* genes were cloned into the broad-host range plasmid pVLT35 behind the IPTG-inducible tac-promoter (Lorenz, V. et al., *Gene* 123: 17-24 (1993)) and electroporated into the *P. putida* mutant. Streptomycin-resistant transformants were subcultured onto minimal medium containing either octanoate or gluconate as sole carbon source. The Nile Blue A fluorescence stain (Page, W. J. and C. J. Tenove, *Biotechnology Techniques* 10: 215-220 (1996)) was used to visualize PHA accumulation. Upon IPTG induction PHA accumulation was observed with pVLT35\_phaC1\_ICL and pVLT35\_phaC2\_ICL, but not with pVLT35 alone, thus indicating that the modified genes were still active.

#### EXAMPLE 3: Plant transformation and screening for PHA synthase C1 transgenic plants

The NotI-cassettes of plasmids pART7\_phaC1\_ICL and pART7\_phaC2\_ICL containing the modified genes flanked by the Cauliflower mosaic virus 35S promoter (CMV35S) and the octopine synthase (ocs) 3'-terminator were cloned into the plant binary vector pART27 to obtain pART27\_phaC1\_ICL and pART27\_phaC2\_ICL. These plasmids were transformed into *A. thaliana* ecotype Columbia by *Agrobacterium* GV3101-mediated transfer utilizing an *in planta* vacuum-infiltration method (Bechtold, N. et al., *C.R. Acad. Sci. Paris* 316: 1194-1199 (1993)). Transgenic T1 plants were selected for antibiotic resistance during germination of the seeds of infiltrated plants on plant growth medium containing mineral salts, sucrose and kanamycin. Negative control plants containing only the insert-less T-DNA of the vector pART27 were obtained in the same way.

Transgenic PHAC1 plants (T1) expressing high amounts of PHA synthase C1 were selected by Western analysis with an antiserum against the PHA synthase C1, which was

obtained from Steinbüchel's laboratory. Unfortunately no antibodies against PHA synthase C2 were found to be suitable, so a different screening strategy was used, see below. Six independent lines expressing varying quantities of PHA synthase C1 were obtained from 12 originally infiltrated plants, which had been harvested individually (another 19 have not yet been investigated). Initially some problems with the western analysis were encountered, one of which was the precipitation of the PHA synthase in plant protein extracts upon freezing. Analysis of the kanamycin segregation of the second generation (T2) and third generation (T3) plants indicated that three of these lines contained multicopy T-DNA inserts. Initially these lines exhibited the highest expression of PHA synthase C1 as judged by western analysis, however, the expression of the transgene in these lines was variable in plants of the T2 and T3 generation and complete "silencing" was observed. The line PHAC1#3.3 was finally chosen for further studies, because it contained a single-locus T-DNA insert and exhibited stable expression of the transgene as seen on the western blot.

#### EXAMPLE 4: PHA production by PHAC1 plants

A protocol for the detection of monomers of PHA by gas chromatography was developed based on the method described for the extraction of PHB from *Arabidopsis* (Poirier, Y. et al., *Int. J. Biol. Macromol.* 17: 7-12 (1995)). Whole leaves were extracted several times with ethanol and methanol to elute all the soluble lipids, thereafter chloroform and methanol acidified with 3% (v/v) H<sub>2</sub>SO<sub>4</sub> were added in equal volumes and the reactions were put at 98°C for 4 hours to transesterify the PHA polyester. GC-chromatograms of the resulting chloroform extracts showed a large number of peaks, most of which were due to the derivatization of various leave compounds. Peaks corresponding to the standards of the expected methyl esters of PHA monomers were, however, distinguishable amongst the others. A large fraction of the plant material was solubilized during this transesterification treatment, it was however not determined whether underivatized PHA remained in the solid underivatized material. This made the quantification of the PHA in plant material slightly uncertain, but the authors estimated intuitively that most of the PHA in the material became derivatized preferentially. The GC-standards (from Sigma Chemical, St. Louis, MO, except H6 which was from Beal Keller) were the methyl esters of D-3-hydroxy-hexanoic acid (3-

OH-caproic acid, H6 monomer), DL-3-hydroxy-octanoic acid (3-OH-caprylic acid, H8 monomer), DL-3-hydroxy-capric acid (H10 monomer), DL-3-hydroxy-lauric acid (H12 monomer) and DL-3-hydroxy-myristic acid (H14 monomer).

The transgenic plants expressing the PHA synthase C1 showed a significant increase in the size of the peaks corresponding to the H6-H14 monomers compared to the negative control plants. One novel peak was found only in PHAC1 plants and never in the negative controls. GC-MS was used to confirm that the peaks observed in both the PHAC1 plants and the negative controls were really identical to the standards and the novel peak was determined as being due to 3-hydroxy-octenyl-methyl-ester containing a single unsaturated bond (H8:1 monomer). It is being speculated that the unsaturated bond is located at carbon 5 and has the cis conformation and that this monomer is due to the degradation of  $\alpha$ -linolenic acid (18:3, all-cis,  $\Delta$ 9,12,15) and 16:3 (all-cis,  $\Delta$ 7, 10, 13) by  $\beta$ -oxidation. This reasoning is based on the prediction, that a D-3-hydroxy-octenyl-CoA  $\beta$ -oxidation intermediate arises due to the cis-double bond at the even-numbered carbons (Gerhardt, B., *Lipid metabolism in plants* (Moore, T. S., Jr., ed.), CRC Press Inc., pp. 527-565 (1993)); see further discussions below under feeding studies). The same argument can be taken for the generation of the other monomers incorporated into the PHA, i.e. that they originated from fatty acids having a double bond at even-numbered carbons, which resulted in the formation of D-3-hydroxy-acyl-CoA  $\beta$ -oxidation intermediates. Thus the H8 monomer would originate from the degradation of linoleic acid (C18:2, all-cis,  $\Delta$ 9,12) or from C16:2, all-cis,  $\Delta$ 7, 10. This however does not satisfactorily explain the whole range of monomers observed, e.g. the H6 monomer would then have to originate from the fatty acids C18:1,  $\Delta$ 14-cis or C16:1,  $\Delta$ 12-cis, while the H14 monomer would have to originate from C18:1,  $\Delta$ 8-cis, or C16:1,  $\Delta$ 4-cis or C14:1,  $\Delta$ 2-cis, etcetera. As most of these would be rather uncommon fatty acids in *A. thaliana*, another argument for the origin of these PHA monomers can be proposed, which is based on the existence of an epimerase activity in plant  $\beta$ -oxidation (Preisig-Müller, R. et al., *J. Biol. Chem.* 269: 20475-20481 (1994)). In this case the D-3-OH-acyl-CoA  $\beta$ -oxidation intermediates are generated at a low rate by the "reverse" reaction catalyzed by the epimerase required for the conversion of D-3-hydroxy-acyl-CoA to the L-form, and sequestration of these D-intermediates into PHA actually

drives the reverse reaction. In this way the whole range of possible monomers can be explained, while the argument involving the unsaturated bond at even-numbered carbons in the acyl chains would still explain the relatively higher proportion of the H8-monomer and the existence of the H8:1 monomer.

Several negative control plants (both *A. thaliana* wild type and pART27 transgenic plants) were analyzed in various experiments without ever seeing more than only trace amounts of the various saturated monomers. The concentrations present in the negative controls were at least 1000 times smaller than in the positive plants, close to the detection limit of the methods at our availability. This was done by utilizing the GC-MS in the SIM mode (selected ion monitoring; ion 103 is characteristic for all of these 3-OH-fatty acid methyl esters) for which the detection limit was found to be approximately 4 pg/ $\mu$ L of the various standards. These compounds in the negative controls might also be intermediates of  $\beta$ -oxidation, i.e. mostly the L-3-hydroxy-acyl-CoAs and perhaps even very low amounts of the D-form, which are normally present at very low concentrations in the plant material in which  $\beta$ -oxidation is taking place. A rough calculation indicated a total PHA content of 0.03% (w/dry weight) in PHAC1#4.4 (multilocus plant), which related to approximately 5  $\mu$ g of PHA in a large fresh leave weighing 155 mg. It was approximated that line PHAC1#3.3 produced 0.01% (weight/dry weight) in soil-grown plants.

#### EXAMPLE 5: Screening for PHA synthase C2 expressing plants

PHAC2 plants were screened directly for PHA production by analysis of dry leaves of T2 plants. Almost all of the T2 plants derived from 13 independently transformed plants were found to produce PHA in varying quantities, as judged by the presence of the novel peak due to the C8:1 monomer and also the peaks of the other PHA monomers. The highest producing plants were analyzed further and homozygous T3 plants were obtained. Two homozygous single-locus T3 lines were selected, PHAC2#19.5 and PHAC2#8.6. In comparison to PHAC1#3.3 plants, these PHAC2 plants produced slightly smaller quantities of PHA in seedlings grown on plates containing MS salts, kanamycin and sucrose. The

monomer composition of the respective transgenic plants was however identical. For that reason most of the further studies were only done with line PHACI#3.3.

#### EXAMPLE 6: Immunolocalization and observation of PHA granules

For the immunolocalization of the peroxisomally-targeted PHA synthase C1, T3 seedlings of lines PHACI#3.3 and PART27#21 (negative control) were grown on plates containing MS salts, kanamycin and sucrose. Seedlings were grown for 7 days under continuous light or in the dark after one day of illumination, the latter was done to obtain etiolated seedlings in which glyoxysomes are more abundant. The seedlings were fixed and sent together with some anti-PHA synthase C1 antiserum to Prof. Leech's laboratory at the University of York, where the immunolocalization was performed. It was found that the peroxisomes in PHACI seedlings were initially difficult to identify, since they did not look normal due to the presence of granules within them. These granules were very abundant in the etiolated seedlings, while in the light-grown seedlings most of the peroxisomes still looked normal or seemed to contain only tiny granules. The PHA synthase C1 was located in what seem to be two different types of organelles or peroxisomes, because the one contains a large quantity of PHA granules while the other contains apparently none. The darker peroxisomes without granules corresponded in appearance most closely to the normal peroxisomes in the negative controls. It is possible that this apparent heterogeneity is simply the results of non-homogenous distribution of granules within the peroxisomes. Glycolate oxidase was used as marker enzyme for peroxisomes of seedlings grown under light, while rubisco was used as chloroplastic marker. Antibodies against these two marker enzymes clearly identified the respective organelles in both PHACI seedlings and in the PART27 negative controls. Glycolate oxidase was found to be located in the organelles, i.e. the peroxisomes, containing PHA granules. Similarly the enzyme isocitrate lyase (ICL) was used as glyoxysomal marker in etiolated seedlings and it also confirmed that the granule-containing organelles were glyoxysomes. The antiserum against PHA synthase C1 unambiguously identified the peroxisomal localization of the PHA synthase in the PHACI seedlings, while it did not detect anything in the negative controls. Unusual accumulations of granules were also observed occasionally in the vacuoles of etiolated PHACI seedlings

and these globules were gold-labelled with anti-PHA synthase C1. This was in correspondence with the observation that the PHB synthase is found on the surface of PHB granules in bacteria (Gengross, T. U. et al., *J. Bacteriol.* 175, 5289-5293 (1993)).

#### EXAMPLE 7: Changing PHA yield and monomer composition in feeding studies

Line PHACI#3.3 was used to investigate if the total yield of PHA could be increased or if PHAs containing other monomers than the "native" PHA could be synthesized in PHACI transgenic plants. For that purpose seeds were sterilized and germinated in liquid medium containing mineral salts and 2% (w/v) sucrose supplemented with fatty acids or other compounds known to be degraded by  $\beta$ -oxidation. In experiment #1 the seedlings were grown for 3 days in the light before the substrates were added and the plant were moved into the dark. The material was harvested after 8 days and derivatized samples were analyzed by gas chromatography.

The results summarized in Table 1 point out several encouraging aspects. The yield of native PHA (obtained without feeding any substrate) was doubled when seedlings were germinated in the dark as opposed to continuous illumination. This could perhaps be ascribed to a more complete mobilization of the seed lipids in etiolated seedlings. In this respect the regulation of the glyoxylate cycle enzymes malate synthase and isocitrate lyase might play a role by affecting lipid-mobilization via  $\beta$ -oxidation. It has been shown that these glyoxylate cycle enzymes are regulated transcriptionally by three types of signal, namely light regulation, carbon catabolite repression by various sugars and developmental regulation during germination and senescence (Graham, I. A. et al., *Plant Mol. Biol.* 15: 539-549 (1990); Graham, I. A. et al., *Plant Cell* 4:349-357 (1992); Graham, I. A. et al., *Plant Cell* 6: 761-772 (1994)).

The large increase in the PHA yield obtained by the feeding of TWEEN-20 (Sigma, 50% palmitic acid (C16) esterified with polyoxyethylenesorbitol, the remainder is made up by lauric acid (C12) and myristic acid (C14) also esterified) (TWEEN is a registered trademark of ICI Americas, Inc., Wilmington, DE) indicated that the PHA synthase was

very active in these plants and thus not responsible for the relatively low yield of native PHA in seedlings grown without added fatty acids. The most pronounced effect of TWEEN-20 on the monomer composition was the decrease in the content of the H8:1 monomer from about 30% in native PHA to about 1%, which was most likely due to the lack of unsaturated fatty acid derivatives in the TWEEN-20. The relative distribution of the other monomers could be explained by the step-by-step  $\beta$ -oxidation of the C16, C14 and C12 components in TWEEN-20. A negative effect on seedling growth due to TWEEN-20 was observed, but it was small considering its high concentration (5% v/v) in the medium.

The accumulation of PHA granules in PHAC1 seedlings grown in liquid cultures supplemented with 5% TWEEN-20 under constant illumination for 12 days was very striking on electron microscope micrographs. These PHA granules were not observed in the negative controls, i.e. pART27 transgenic seedlings fed with TWEEN-20. The granules looked different from the starch granules observed in chloroplasts. These electron microscopic studies were done in our own institute by Mrs J. Petétot and the results confirmed similar results obtained with etiolated seedlings in Prof. Leech's laboratory.

TWEEN-60 (Sigma; 50% stearic acid (C18) and some palmitic and myristic acid; all esterified to polyoxyethylenesorbitol) and TWEEN-80 (Sigma; 50% oleic acid (C18:1), esterified to polyoxyethylenesorbitol) had less impact on the PHA yield, the monomer composition and the seedling growth than TWEEN-20. The relatively high level of the H8:1 monomer might be due to a higher contamination of TWEEN-60 and -80 with unsaturated fatty acids like  $\alpha$ -linolenic acid, see above.

The free fatty acids hexanoate and octanoate were fed at very low concentrations due to their toxic effects on plant growth. For hexanoate a large increase of the H6 monomer was observed, while octanoate resulted in a very high increase of the H8 monomer together with a moderate increase in the H6 monomer. For both substrates the H8:1 monomer content remained relatively high, which was probably due to the normal accumulation of PHA from endogenous lipid  $\beta$ -oxidation ("native" PHA).

Table 1. Increasing the total yield of PHA and changing its monomer composition in PHAC1 seedlings germinated in liquid media supplemented with fatty acids

Substrate	[sub] % light or dark (day weight in 4 to 12)	fresh weight in g x 100 <sup>b</sup>	mg PHA <sup>a</sup> per g fresh weight x 100 <sup>c</sup>	% of total PHA (w/w)							
				H6	H8	H8:1	H10	H12	H14		
None	light	232	1.9	1.3	43	29	10	9.7	6.8		
None <sup>c</sup>	dark	186 $\pm$ 25	4.4	1.4	42	32	9.2	9.2	6.4		
TWEEN-20	light	142	64	3.8	37	1.1	27	28	3.4		
TWEEN-20 <sup>c</sup>	dark	57 $\pm$ 24	70	4.0	41	1.5	25	25	2.9		
TWEEN-60 <sup>c</sup>	dark	125 $\pm$ 55	9.7	2.2	37	17	16	18	10		
TWEEN-80 <sup>c</sup>	dark	141 $\pm$ 34	6.5	3.2	44	20	15	14	4.0		
Hexanoate (C6) <sup>c</sup>	dark	70 $\pm$ 3	11	30	32	21	7.0	7.4	3.1		
Octanoate (C8) <sup>c</sup>	dark	125 $\pm$ 44	16	5.2	73	13	3.7	3.7	1.5		

<sup>a</sup> The transesterified plant material (of specified weight) was in a volume of 1 mL chloroform, of which 1  $\mu$ L was analyzed by GC.

<sup>b</sup> An average of 30 seedlings were grown per sample.

<sup>c</sup> Samples were done in duplicate and the results were averaged.

In experiment #2 (Tables 2 and 3) the seedlings were germinated for 8 days under continuous illumination, then the growth medium was replaced by the same medium containing 5% (v/v) TWEEN-80 together with various fatty acids, the purpose of the TWEEN-80 was to solubilize the water-insoluble fatty acids. The samples were placed back under constant illumination for another 6 days before being harvested and analysed. All samples were done in duplicate and each sample contained approximately 45 seeds which were germinated together in a large capped test-tube. Negative controls with pART27 plants were done for each substrate in the identical fashion. None of the novel PHA-monomer peaks were found in these negative controls.

Feeding of the saturated fatty acid tridecanoic acid (C13) and the branched fatty acid 8-methyl-nonanoic acid (8M-C9) resulted in the incorporation of a whole range of novel monomers. The identity of all these novel monomers was established by GC-MS. All of them had an uneven number of carbon atoms in their acyl chains and could be directly

traced to the original fatty acid supplied in the medium or intermediates of its degradation by  $\beta$ -oxidation. For tridecanoic acid, transgenic PHAC1 plants were found to contain a polymer having H13-, H11-, H9- and H7-3-hydroxy-alkanoic acid monomers. In the case of 8M-C9 the two novel monomers, 8-methyl-3-D-hydroxy-nonanoic acid (8M-H9) and 6-methyl-3-D-hydroxy-heptanoic acid (6M-H7), retained the branched structure of the original substrate. This shows that the PHA synthase C1 was able to incorporate a large variety of monomers into the polymer, provided that intermediates having the proper conformation were generated. The descending order in terms of quantities of the novel monomers (H13>H11>H9>H7; and 8M-H9>6M-H7) suggests that the  $\beta$ -oxidation of these unusual fatty acids proceeds slowly, thus permitting more time for intermediate-sequestration by the PHA synthase. It is possible that the 3-hydroxy-acyl-CoA dehydrogenase (MFP) and some other enzymes of the  $\beta$ -oxidation cycle have a low substrate specificity for these fatty acids and their derived intermediates.

Feeding of petroselinic acid (C18:1, 6-cis) resulted in a large increase in the content of the H14 monomer. This observation was in agreement with the proposed scheme of its degradation by  $\beta$ -oxidation (Gerhardt, B., *Lipid metabolism in plants* (Moore, T. S., Jr., ed.), CRC Press Inc., pp. 527-565 (1993)). All unsaturated bonds in the cis-conformation starting at an even-numbered carbon in the acyl chain were proposed to present obstacles to the normal cycle of the  $\beta$ -oxidation and had to be circumvented by modifications of the pathway. This is because the D-3-hydroxy-acyl-CoA can be formed by the action of the enoyl-CoA hydratase (MFP) from 2-cis-enoyl-CoA (cis-unsaturated bond in even-numbered position), but the D-3-hydroxy-acyl-CoA cannot be utilized by the 3-hydroxy-acyl-CoA dehydrogenase (MFP), which can only act on the L-3-hydroxy-acyl-CoA. Three possible modifications were put forward: 1) An epimerase converts the D-3-hydroxy-acyl-CoA to the L-form. 2) A dehydratase (also called D-3-hydroxyacyl-CoA hydrolase or D-specific 2-trans-enoyl-CoA hydratase II, see Engeland, K. and Kindl, H., *Eur. J. Biochem.* 200: 171-178 (1991)) converts the D-3-hydroxy-acyl-CoA to 2-trans-enoyl-CoA, which can then be reconverted to L-3-hydroxy-acyl-CoA by the enoyl-CoA hydratase I. 3) A 2,4-dienoyl-CoA reductase reduces the 2-trans-4-cis-acyl-CoA  $\beta$ -oxidation intermediate to the 3-cis-enoyl-CoA, which in turn will require the activity of an isomerase to form the 2-trans-enoyl-

CoA  $\beta$ -oxidation intermediate. The first two options would result in the generation of D-3-hydroxy-acyl-CoA intermediates which would be directly available to the PHA synthase. Thus the observation of the specific increase in the H14 monomer upon feeding with petroselinic acid fits well with the predicted modifications of the  $\beta$ -oxidation to bypass the cis-unsaturated bond at carbon 6 of petroselinic acid. The same modifications have also been used above to explain the presence of the 3-hydroxy-octenoyl monomer (H8:1) in the native PHA. It was speculated that this monomer was due to the degradation of 18:3, all-cis- $\Delta$ 9, 12, 15 and 16:3, all-cis- $\Delta$ 7, 10, 13 by  $\beta$ -oxidation. The high proportion of H8 monomer could similarly be due to the degradation of linoleic acid (18:2, all-cis- $\Delta$ 9,12) which is an abundant fatty acid in plant material.

The degradation of fatty acids containing hydroxy groups on even-numbered carbon atoms in either the D- or the L-conformation also poses obstacles to the normal  $\beta$ -oxidation pathway and modifications are required to bypass these (Gerhardt, B., *Lipid metabolism in plants* (Moore, T. S., Jr., ed.), CRC Press Inc., pp. 527-565 (1993)). The D-4-hydroxy-decanoate-CoA and D-2-hydroxy-octanoate-CoA intermediates were predicted to arise in the degradation of ricinoleic acid (D-12-hydroxy-oleic acid (9-cis)). To investigate whether these intermediates might be incorporated into the PHA polymer by the PHA synthase, ricinoleic acid was used to supplement the medium in which PHAC1 plants were germinating. No major peaks due to the incorporation of novel monomers into the PHA polymer were detected, but GC-MS analysis was utilized to search for specific predicted novel monomers by looking for characteristic fragmentation products, namely ions 117 and 89. A small peak was found with ion 117, this peak showed the fragmentation fingerprint of the D-4-hydroxy-decanoate-methyl ester and was absent in the corresponding negative control. No novel peak was found with ion 89, thus excluding the possibility that the D-2-hydroxy-octanoate was incorporated into the polymer. It is known that the PHA synthase can incorporate D-4-hydroxy- and D5-hydroxy monomers into PHA in bacterial cultures, therefore the incorporation of the D-4-hydroxy-decanoate in the germinating seeds fed with ricinoleic acid was plausible. The very low abundance of the monomer could perhaps be explained by an alternative and more efficient pathway for the degradation of ricinoleic



(Gerhardt, B., *Lipid metabolism in plants* (Moore, T. S., Jr., ed.), CRC Press Inc., pp. 527-565 (1993)).

Table 2. Quantity of PHA production in PHAC1 seedlings germinated in liquid medium supplemented with fatty acids

Substrate	[Substrate] % (w/v)	fresh weight (mg)	mg PHA <sup>a</sup> per gram fresh weight x 0.01
None		458 ± 8	4.6
5% TWEEN-80 (T)		549 ± 43	5.0
Tridecanoic acid (C13) + T	0.1	276 ± 9	28
8-methyl-nonanoic acid (8M-C9) + T	0.1	48 ± 14	46
Petroselinic acid (C18:1, 6-cis) + T	1	287	9.4
Ricinoleic acid (D12-OH-C18:1, 9-cis) + T	0.1	215 ± 21	6.0

<sup>a</sup> The plant material (of specified weight) was transesterified in different volumes, but the integrated peak-areas were calculated to homologate the sample-volumes (1 mL chloroform, of which 1 µL was analyzed by GC).

Table 3. Monomer composition of PHA produced in PHAC1 seedlings germinated in liquid medium supplemented with fatty acids

Substrate	% of total PHA (w/w)											
	H6	H7	6M- H7 <sup>a</sup>	H8	H8:1	H9	8M- H9 <sup>b</sup>	H10	4OH-H11 H10 <sup>c</sup>	H12	H13	H14
None	1.0			36	28			10		14		10
5% TWEEN 80 (T)	2.1			44	14			19		15		6.3
Tridecanoic acid (C13) + T	0.46	9.6		9.1	2.4	18		3.4		20	3.2	32
8-methyl- nonanoic acid (8M-C9) + T	0.20		24	9.2	3.0		55	3.0		2.8		2.2
Petroselinic acid (C18:1 6-cis) + T	1.6			32	7.4			17		17		26
Ricinoleic acid (D12 OH-C18:1, 9 cis) + T	1.8			46	8.8			30	1.2 <sup>c</sup>	7.4		5.0

<sup>a</sup> 8M-H9 and 6M-H7 refer to 8-methyl-3-D-hydroxy-nonanoic acid and 6-methyl-3-D-hydroxy-heptanoic acid, respectively.

<sup>b</sup> 4-OH-H10 refers to D-4-hydroxy-decanoate.

<sup>c</sup> The quantity of 4-OH-H10 was estimated by comparing peak sizes with H6 on a GC-MS chromatogram.

#### EXAMPLE 8: Extraction of high molecular weight PHA

The presence of derivatized monomers of PHA in PHAC1 plants had been established by the GC-analysis of trans-esterified intact plant material. To prove that the PHA was synthesized as high-molecular weight polymer and for its physico-chemical characterization, the purification of large quantities (i.e. in the mg range) was undertaken. Seeds of PHAC1#3.3 were germinated in liquid medium with and without addition of TWEEN-20 in order to obtain TWEEN-20-derived PHA or unmodified PHA, respectively.

For the TWEEN-20-derived PHA, approximately 16000 seeds (313 mg dry seeds) were germinated in 900 mL 1/2xMS + 1% sucrose medium for 7 days under continuous illumination on a shaker, the medium was replaced with 1/2xMS + 2% sucrose containing 5% TWEEN-20 and the seedlings were grown for another 9 days in the light. The plant material was harvested, washed extensively with water to remove residual TWEEN-20, frozen and lyophilized. The dry material was ground with a mortar and pestle, weighed, and lipids were extracted by a six-hour Soxhlet-extraction with methanol. The methanol-insoluble PHA was extracted for 24 hours in the same manner with chloroform. The chloroform extract was concentrated under reduced pressure and the PHA was precipitated by the addition of 10 volumes of cold methanol. This methanol precipitation was performed twice to ensure a high purity of the PHA. 27 mg of PHA was thus obtained from 5.35 g lyophilized and powdered seedling material, which related to 0.50% weight/dry weight. The PHA was trans-esterified and analyzed by GC. It was found that 58% of the PHA present in the methanol-extracted plant powder was extracted by the chloroform. It has been established in previous experiments that this remaining PHA was recalcitrant to extraction. The chromatogram showed that the extracted PHA was adequately pure with the peaks of the six identified monomers constituting 93% of the total integrated area. The ratio of the integrated areas between the different monomers was very similar to the result shown in Table 1 for the sample containing TWEEN-20 and grown under light, see Table 4.

For the extraction of high-M<sub>n</sub> PHA produced by PHA<sub>1</sub> plants without additional fatty acid supplements (native PHA), 1076 mg seeds (approx. 54000 seeds) were germinated in 3.3 L liquid medium (1/2xMS, 2% sucrose). The seeds were germinated under continuous illumination for 6 days, thereafter the medium was replaced and the seedlings put into the dark for another 7 days in order to induce plant senescence. The PHA was extracted from the plant material as above and one methanol precipitation was performed to purify the PHA. 23 mg of PHA was obtained from 14.3 g dry plant material, which related to 0.16 % weight/dry weight. It was determined that ≥69 % of the PHA had remained in the plant material after the chloroform extraction, which could be due to either the high content of C8:1 monomer (see Table 5) causing the polymer to "stick", or due to moisture in the ground material which had not been lyophilized completely, or due to the

large sample size for which a longer and more efficient chloroform extraction might have been required. The purification of native PHA and analysis by GC-MS allowed the detection of several more peaks that could not be initially resolved in crude extracts because of the high level of noise in the chromatogram. A total of eighteen 3-hydroxyacid monomers could be detected in the polymer (Table 1). In addition to 3-hydroxyhexanoic acid (H:6), 3-hydroxyoctanoic acid (H:8), 3-hydroxydecanoic acid (H:10), 3-hydroxydodecanoic acid (H:12), 3-hydroxytetradecanoic acid (H:14) and 3-hydroxyoctenoic acid (H:8:1) monomers previously detected in the transesterification of intact plant material (crude extract) (Table 1), novel saturated and unsaturated monomers were detected which include 3-hydroxyhexadecanoic acid (H:16), 3-hydroxynonanoic acid (H:9), 3-hydroxyundecanoic acid (H:11), 3-hydroxytridecanoic acid (H:13), 3-hydroxyhexadecatrienoic acid (H:16:3), 3-hydroxyhexadecadienoic acid (H:16:2), 3-hydroxyhexadecenoic acid (H:16:1), 3-hydroxytetradecatrienoic acid (H:14:3), 3-hydroxytetradecadienoic acid (H:14:2), 3-hydroxytetradecenoic acid (H:14:1), 3-hydroxydodecadienoic acid (H:12:2) and 3-hydroxydodecenoic acid (H:12:1). All even-chained monomers could be quantified and results are shown in Table 5.

It is expected that many of the unidentified minor peaks detected in the PHA purified from the TWEEN-20-fed seedlings would correspond to some of the minor saturated and unsaturated monomer detected in the "native" PHA.

Table 4. Comparison of the monomer composition of purified high-molecular weight PHA from TWEEN-20 fed plants with results obtained for transesterified intact seedlings during the preliminary feeding studies

Sample	H:6	H:8	H:8:1	H:10	H:12	H:14
Purified high M <sub>n</sub> TWEEN-20 derived PHA	1.9	33.5	0.50	29	32	2.7
TWEEN-20 + light see Table 1, line 3	3.8	37	1.1	27	28	3.

\* Integrated area on the chromatogram.

Table 5. Monomer composition of "native" PHA isolated from *phaC1*-transformed plant line 3.3 grown in liquid media<sup>a</sup>

Monomer	H6	H8	H8:1	H10	H12	H12:1	H12:2
% (w/w)	1.1	23	18	4.7	5.8	4.3	5.0
std. dev.	0.16	4.4	4.6	0.51	0.46	0.60	1.3
Monomer	H14	H14:1	H14:2	H14:3	H16	H16:2	H16:3
% (w/w)	4.2	6.7	7.5	11	2.0	2.0	5.6
std. dev.	1.1	2.3	1.4	3.2	0.26	0.41	1.3

<sup>a</sup> Quantification of methyl esters was performed with a GC with a FID detector. Values were obtained from four separate PHA preparations. Monomers present in trace amounts (H9, H:11, H:13, H16:1) were not quantified.

#### EXAMPLE 9: Chemical characterization of high-molecular weight plant PHA

Purified TWEEN-20-derived PHA (13 mg) and unmodified PHA (5 mg) were given to Géraldine Coullerez at the EPFL (collaboration IBPV-EPFL) for the physico-chemical characterization of the polymer. Two different samples of bacterial PHA, PHA1 and PHOE, were obtained from Witholt and Kellerhals (ETH Zürich) to be used as controls. PHA1 contained predominantly H6 and H8 monomers (10% and 90%, respectively), while PHOE contained 4-10% H8:1, the balance being H6 and H8. The molecular weights and the respective dispersion coefficients of the polymers were determined by gel permeation chromatography (see Table 6). Polystyrene polymers were used as molecular weight standards. The results clearly show that the TWEEN-20 derived PHA produced by the transgenic plants is in the form of a high-M<sub>r</sub> polymer (about 200-250 monomers), although the molecular weight is only 20-25% of the bacterial polymers (about 1000 monomers). This shorter polymer length can be explained by an overabundance of PHA synthase relative to its substrate concentration and similar results have also been obtained in *in vitro* polymerization assays with purified PHB synthase (Jun Sim, S. et al., *Nature Biotechnology* 15: 63-67 (1997)). It is also possible that PHA polymers with longer chain lengths are trapped in the plant material, since a significant proportion of the PHA seems to be

recalcitrant to chloroform extraction ( $\geq 50\%$ , difficult to determine exact amounts in the trans-esterification of intact or powdered plant material, see above).

NMR analysis of the plant and bacterial PHAs revealed, that the TWEEN-20 derived plant PHA had the same structure as the bacterial PHA. The NMR spectrum of the unmodified plant PHA showed the peaks characteristic for the PHA polymer backbone, as well as several other peaks which have not been properly assigned or identified at this stage, but which could be due to various unsaturated bonds in the side chains of the polymer.

Table 6. Comparison of molecular weights of high-M<sub>r</sub> PHA<sub>mod</sub> purified from plants and bacteria

Origin of PHA	Mw	Mn	Dispersion Index
TWEEN-20 derived PHA - <i>Arabidopsis</i> PHAC1#3.3	$4.01 \times 10^4$	4910	8.17
PHA1 from <i>Pseudomonas oleovorans</i>	$1.46 \times 10^5$	58850	2.48
PHOE from <i>P. oleovorans</i>	$2.0 \times 10^5$	81590	2.44

EXAMPLE 10: The multifunctional protein (MFP) from the yeast *Candida tropicalis*

In animals, plants and bacteria,  $\beta$ -oxidation has been shown to proceed via the L-isomer of the 3-hydroxy-acyl-CoA intermediates and any D-isomers (which are predicted to arise in the degradation of fatty acids containing cis-unsaturated bonds at even-numbered carbons) have to be converted to the L-form in order to be oxidized further by the dehydrogenase activity of the multifunctional protein (MFP). In yeast the  $\beta$ -oxidation was reported to proceed via the D-isomer (Nuttley, W. M. et al., Gene 69: 171-180 (1988); Hiltunen, J. K. et al., *J. Biol. Chem.* 267: 6646-6653 (1992); Fossá, A. et al., *Mol. Gen. Genet.* 247: 95-104 (1995)). The yeast multifunctional protein (MFP) was shown to contain enoyl-CoA hydratase II and D-3-hydroxyacyl-CoA dehydrogenase activities, which together converted trans-2-enoyl-CoA via D-3-hydroxyacyl-CoA to 3-ketoacyl-CoA, i.e. the D-isomer was directly utilized by the dehydrogenase without prior conversion to the L-form.

It is anticipated that expression of this hydratase II activity together with the PHA synthase in the peroxisomes of double-transgenic plants will generate more of the D-3-hydroxy-acyl-CoA intermediates for their incorporation by the PHA synthase into the PHA polymer, thus increasing the final yield of PHA. Four separate approaches are envisioned.

#### A. Expression of the unchanged MFP from *C. tropicalis* in *A. thaliana*.

Since the hydratase II activity forms part of the MFP it was decided to perform investigatory experiments with the complete MFP prior to attempting to abolish the D-3-hydroxyacyl-CoA dehydrogenase activity. As the fungal MFP already had a peroxisomal targeting signal, this protein was expected also to be targeted to the plant peroxisomes.

The *C. tropicalis* MFP cDNA (Nuttley, W. M. et al, *Gene* 69: 171-180 (1988), GenBank Accession Number M22765) was cloned via PCR amplification (SEQ ID NO:21, encoding SEQ ID NO:22) into pART7 to obtain pART7\_MFP. The NotI-cassette, containing the CAMV35S-promoter in front of the MFP gene and the ocs3'-terminator, was inserted into the plant binary vector pART27 to obtain pART27\_MFP, which was transformed into *Arabidopsis*. Transgenic plant were selected on kanamycin and screened for the expression of the MFP protein with an anti-MFP antiserum. Homozygous T2 plants were cross-fertilized with PHAC1#3, PHAC1#4 and PHAC1#9 plants. Offspring from these crosses will be analyzed for their ability to biosynthesize PHA.

#### B. Changing the peroxisomal targeting signal of the yeast multifunctional protein (MFP) from -AKI to -SKL.

The COOH-terminal tripeptide -AKI was shown to be responsible for peroxisomal targeting of the MFP in yeast, but it has not yet been demonstrated to function in plant peroxisomal targeting. The MFP SKL gene, in which the 3'-terminal nucleotide sequence of the MFP gene encoding the -AKI tripeptide had been changed to -SKL by PCR site-directed mutagenesis (SEQ ID NO:23, encoding SEQ ID NO:24), was obtained from the laboratory of K. Hillman to ascertain that the MFP was properly targeted to the plant peroxisomes and

to serve as a positive control in targeting studies with the yeast multifunctional protein (MFP) in plant cells. The MFP SKL gene was used to construct pART7\_MFP SKL. The NotI-cassette of pART7\_MFP SKL, containing the MFP SKL gene flanked by the CAMV35S promoter and the ocs3'-terminator, was cloned into pART27 to obtain pART27\_MFP SKL, which was transformed into *A. thaliana* ecotype Columbia. Kanamycin resistant T1 plants were obtained. The high-MFP SKL-expressing lines will be selected by Western analysis of T2 plants, and the selected lines will be crossed with PHAC1#3.3 plants.

#### C. Deleting the peroxisomal targeting signal of the yeast multifunctional protein (MFP).

The construct pART7\_MFPΔAKI was obtained by PCR amplification of the MFP gene such that the 3'-terminal nucleotide sequence of the MFP gene encoding the -AKI tripeptide was deleted by the introduction of a stop codon (SEQ ID NO:25, encoding SEQ ID NO:26). The "detargeted" MFPΔAKI is expected to be localized in the cytoplasm and will be utilized as negative control in experiments to study the localization of MFP and MFP SKL in plant cells. pART27\_MFPΔAKI was transformed into *A. thaliana* ecotype Columbia and Kanamycin resistant T1 plants were obtained. The high-MFPΔAKI-expressing lines will be selected by Western analysis of T2 plants and these lines will be crossed with PHAC1#3.3 plants.

#### D. Deleting the dehydrogenase activity of the yeast multifunctional protein (MFP).

As only the hydratase II activity of the yeast multifunctional protein (MFP) is of interest, plants will be transformed with the MFPΔDH gene, in which the dehydrogenase activity was deleted by site-directed mutagenesis of specific amino acid residues identified as being essential for this activity.

#### EXAMPLE II: Verification of enzyme activity of modified MFP constructs in *Pichia*

The modified MFP.SKL and MFPΔAKI genes were subcloned from pART7\_MFP.SKL and pART7\_MFPΔAKI into the yeast expression vector pHILD2. The resulting plasmids pHILD2\_MFP.SKL and pHILD2\_MFPΔAKI were transformed into *Pichia* and enzyme assays were performed in Hiltunen's laboratory. Results indicated that the modifications to the genes did not have an effect on the dehydrogenase and the hydratase enzymatic activities.

EXAMPLE 12: Expression of the FatB3 acyl-ACP thioesterase in double transgenics to increase PHA yield

Expression of the California bay acyl-ACP thioesterase was shown to cause premature termination of fatty acid elongation during fatty acid biosynthesis in transgenic oilseed plants (Vocler, T. A. et al., *Science* 257: 72-74 (1992)). The resulting medium-chain-length fatty acids were found to accumulate in the triglycerides of seed lipids, but could not be detected in leaves. It is thought that medium chain fatty acids do not accumulate in the leaves of transgenic plants because they get degraded immediately by  $\beta$ -oxidation (Eccleston, V. S. et al., *Planta* 198: 46-53 (1996)). This increased flux of medium-chain fatty acids through  $\beta$ -oxidation may be exploited to improve the yield of PHA, as well as to modify the composition of the polymer towards saturated H6-H14 monomers in double transgenic plants expressing both acyl-ACP thioesterase and the PHAC1 synthase.

The plasmid pB149\_FatB3 containing the *Cuphea lanceolata* thioesterase FatB3 gene under control of a 200 bp minimal promoter derived from the 35S promoter was infiltrated into the *A. thaliana* PHAC1#3.3 transgenic line which is homozygous for the PHAC1 gene. Hygromycin resistant lines were obtained and the seed lipid content of T1 seeds was analysed for increased levels of medium chain length fatty acids and 11 separate lines expressing high levels of the acyl-ACP thioesterase were identified in this manner. Subsequently the polyhydroxyalkanoate content of leaves from soil grown T2 double transgenic offspring was determined by GC and GC-MS analysis of the 3-hydroxy-fatty

acid methyl esters obtained by transesterification of whole leaves. The results (Table 7) indicated an approximate tenfold increase in the polyhydroxyalkanoate content of leaves from double transgenic plants when compared to plants expressing only the PHAC1 synthase. The increased polyhydroxyalkanoate yield was mainly due to a large increase in the content of the saturated polyhydroxyalkanoate monomers with an even number of carbons, namely 3-OH-octanoate (H8), 3-OH-decanoate (H10), 3-OH-dodecanoate (H12) and 3-OH-tetradecanoate (H14) (Table 8).

The recombinant FatB3 acyl-ACP thioesterase is naturally targeted to the chloroplast, where it removes medium chain-length acyl-ACP intermediates from the fatty acid biosynthesis. These short chain fatty acids accumulate in the seed lipids, but not in the leaves of transgenic plants and it has been speculated, that they are immediately degraded by  $\beta$ -oxidation. Results with these double transgenic plants indicate that there is indeed an increase in the  $\beta$ -oxidation of medium chain length fatty acids in the leaves, which results in a higher yield of polyhydroxyalkanoate due to the incorporation of the  $\beta$ -oxidation intermediates into the PHA by the polyhydroxyalkanoate synthase.

Table 7. PHA content of leaves from single and double transgenic plants expressing the PHAC1 synthase alone or together with the FatB3 acyl-ACP thioesterase

Plants	PHA content (mg/g fresh weight)	
	average	std. deviation
PHAC1#3.3 plant 1	0.0040	
PHAC1#3.3 plant 2	0.0253	0.0147
PHAC1#3.3 + FatB3 line 2.4a plant 2	0.1281	0.015
PHAC1#3.3 + FatB3 line 2.4b plant 1	0.0749	0.1175
PHAC1#3.3 + FatB3 line 2.4b plant 5	0.1495	0.038

Table 8. PHA content of leaves from single and double transgenic plants expressing the PHAC1 synthase alone or together with the FabB3 acyl-ACP thioesterase

Monomer	PHAC1#33	std. dev.	% (w/v)	mg/g	std. dev.	% (w/v)
H6	0.00035	0.00036	2.39	0.00455	0.00131	3.87
H8:1	0.00451	0.00525	30.76	0.00790	0.00913	6.73
H8	0.00205	0.00201	13.94	0.03765	0.01120	32.05
H9	0.00014	0.00001	0.95	0.00029	0.00010	0.25
H10	0.00087	0.00080	5.91	0.04694	0.01816	39.96
H11	0.00017	0.00002	1.17	0.00034	0.00015	0.29
H12	0.00145	0.00145	9.87	0.00642	0.00247	5.47
H13	0.00016	0.00010	1.09	0.00023	0.00013	0.20
H14:1	0.00072	0.00059	4.92	0.00141	0.00114	1.20
H14:2	0.00121	0.00142	8.21	0.00179	0.00209	1.53
H14:3	0.00086	0.00106	5.86	0.00142	0.00178	1.20
H14	0.00219	0.00222	14.93	0.00853	0.00459	7.26

#### EXAMPLE 13: Crossing PHAC1#3.3 transgenic plants with fatty acyl hydroxylase

##### L.FabH2 transgenic plants

Three lines of transgenic *A. thaliana* expressing the L.FabH2 fatty acyl hydroxylase gene from *Lesquerella* were obtained from Pierre Broun (Chris Somerville's laboratory, Carnegie Institution, Stanford, CA). This fatty acyl hydroxylase is responsible for the production of ricinoleic acid (C18:1; 9-cis, D-12-hydroxy) in *Lesquerella*. It was found that hydroxylated fatty acids accumulated in the seed triglycerides of *Arabidopsis*, but not in the leaves, again indicating that hydroxylated fatty acids synthesized in leaves are most likely degraded by  $\beta$ -oxidation (Broun, P. and Somerville, C., *Plant Physiol.* 113: 933-942 (1997); van de Loo, F.N. et al., *Proc. Natl. Acad. Sci. U.S.A.* 92: 6743-6747 (1995)). Crosses were made with the three fatty acyl hydroxylase transgenic lines and the PHAC1#3.3 line and the seeds of these crosses were harvested. Seeds and their progeny plants will be examined for their levels of PHA biosynthesis. The aim of this experiment is to investigate if the increased flux of hydroxylated fatty acids to the  $\beta$ -oxidation cycle in transgenic plants expressing the FabH2 and PHA synthase genes can lead to an increase in the yield of PHA and if novel hydroxylated monomers can be incorporated in the PHA.

#### EXAMPLE 14: Influence of carbon source and light conditions on PHA synthesis

The amount of PHA present in plant tissues was influenced by the growth conditions. For plants grown for three weeks under constant illumination in MS liquid media with 2% sucrose, the yield of PHA was approximately 0.6 mg/g dry weight (dwt). Removal of sucrose for the last week of growth in the light resulted in a 100% increase in PHA, while plants growing in 2% sucrose but shifted in the dark for the last week accumulated 22% more PHA (Table 9).

Table 9. Influence of sucrose and light on PHA accumulation in phaC1-transformed line 3.3

PHA yield	Growth conditions <sup>a</sup>					
	0% sucrose	0.2% sucrose	2% sucrose	0% sucrose	0.2% sucrose	2% sucrose
mg PHA/g dwt	1.42	1.31	0.73	1.23	1.08	0.60
Relative % <sup>b</sup>	100	92	52	87	76	42

<sup>a</sup> Seedlings were grown under constant illumination in a liquid medium containing MS salts and 2% (w/v) sucrose for 2 weeks, and then grown for another week, either in the dark or in the light, in media containing different concentrations of sucrose.

<sup>b</sup> The yield of 1.42 mg/g dry weight was arbitrarily defined as 100%.

#### EXAMPLE 15: Peroxisome targeting

It has been shown in multiple systems (e.g., yeast, animal, and plants) that targeting of proteins to the peroxisome can be achieved by the addition of as little as three amino acids at the carboxy end of a foreign protein (see Gietl, C., *Physiol. Plant.* 97: 599-608 (1996); Purdue, P.E. and Lazarow, P. B., *J. Biol. Chem.* 269: 30065-30068 (1994); Subramani, *Ann. Rev. Cell Biol.* 9: 445-478 (1993)). The minimal consensus sequence for peroxisome targeting of protein via the carboxy end, named PTS1 for peroxisomal targeting sequence 1, is a small uncharged amino acid at position 1 (S, A, or P), a positively-charged

amino acids at position 2 (K, R, S, or H), and a hydrophobic amino acid at position 3 (L, M, I or F).

Thus, although the initial minimal PTS1 sequence was defined as SKL, a range of substitution have been found to be effective PTS1 signal, including ARM, SRM, SKL, ARL, SRL, PSI, or PRM. Specific examples of targeting of foreign proteins in plants include: 6 amino acid PTS1 (RAVARL, Volokita, M., *Plant J.* 1: 361-366 (1991)); 5 amino acids PTS1 (AKSRM, Olsen, L. J. et al., *Plant Cell* 5: 941-952 (1993)); 4 amino acids PTS1 (KSRM, Trelease, R. N. et al., *Protoplasma* 195: 156-167 (1996)); 5 amino acid PTS1 (ELSR), Hayashi, M et al., *Plant J.* 10: 225-234 (1996)); 4 amino acid PST1 (RPSI, Mullen R. T. et al., *Plant J.* 12: 313-322 (1997)); 3 amino acid PTS1 (SKL, Banjoko, A. et al., *Plant Physiol* 107: 1201-1208 (1995)); 3 amino acid PTS1 (ARM, Lee, M.S. et al., *Plant Cell* 8: 185-197 (1997)).

A comparison of the peroxisomal targeting sequence 1 (PTS1) found in mammals, fungi and trypanosomes was performed by Purdue, P.E. and Lazarow, P.B. (*J. Biol. Chem.* 269: 30065-30068 (1994)). All sequences shown in Table 10 are functional in at least one species. Other sequences may or may not have been tested. For trypanosomes, all sequences with a single amino acid change from SKL that are not shown are nonfunctional. The asterisks refer to the fact that -NKL and -SQL (outside the mammalian consensus, but not directly tested) have been found at the C termini of mammalian peroxisomal proteins. Uppercase, functional; lowercase, nonfunctional; underlined, not yet found on a peroxisomal protein in that species.

Table 10. C-terminal peroxisomal targeting sequences.

Mammals	<i>S. cerevisiae</i>	<i>H. polymorpha</i>	<i>C. albicans</i>	Trypanosomes
SKL	SKL	SKL		SKL
SRL				SRL
SHL				SHL
AKL				AKL
CKL				CKL
skf	SKF			
ski		SKI		SKI
*		NKL		NKL
		ARF		
	AKI		AKI	
	agi		AOI	
	gki		GKI	
ssl				SSL
				SKM
tkl				G/H/T-KL
*				S-M/N/O-L
				SKY

The minimal peroxisomal targeting sequence 1 (PTS1) in plants has been found to be ARM, SRM, SKL, ARL, SRL, PSI, and PRM (Compilation from Volokita, M., *Plant J.*, 1: 361-366 (1991); Olsen, L.J. et al., *Plant Cell*, 5: 941-952 (1993); Trelease, R.N. et al., *Protoplasma*, 195: 156-167 (1996); Gietl, C., *Physiol. Plant.*, 97: 599-608 (1996); Purdue, P.E. and Lazarow, P.B., *J. Biol. Chem.*, 269: 30065-30068 (1994); Subramani, *Ann. Rev. Cell Biol.*, 9: 445-478 (1993); Mullen, R.T., et al., *Plant J.*, 12: 313-322 (1997); Lee, M.S., et al., *Plant Cell*, 9: 185-197 (1997)).

Some proteins are targeted to the peroxisome via an N-terminal extension called PTS2 for peroxisome targeting sequence 2. In this case, a consensus sequence of nine amino

acids has been defined, being (R/K)(L/Q/I)XXXXX(H/Q/L). Foreign protein (eg  $\beta$ -glucuronidase) can also be targeted in plants to the peroxisome by adding a PTS2 sequence at the N-terminal end of the protein (Kato et al, *Plant Cell* 8: 1601-1611 (1996)).

#### EXAMPLE 16: Co-expression of PHA with other sequences resulting in increased or novel PHA biosynthesis

PHA<sub>med</sub> synthesized in transgenic plants can include a large variety of monomers, with functional groups that can be used to modify and improve the characteristics of the polymer before or after extraction from the plant. For example, the presence of double bonds, epoxy groups, or acetylated groups within the PHA may be used to cross-link the polymer. The examples herein have demonstrated the incorporation of the following range of monomers into plant PHA<sub>med</sub>: even-chain saturated 3-OH-acyl monomers with six to sixteen carbons; odd-chain saturated 3-OH-acyl monomers with seven to thirteen carbons; unsaturated 3-OH-acyl monomer with 8, 12, 14, and 16 carbons and with 1, 2, or 3 double bonds; branched-chain 3-OH-acyl monomers (8-methyl-3-D-hydroxy-nonanoic acid and 6-methyl-3-D-hydroxy-heptanoic acid) and 4-OH-acyl monomers (D-4-hydroxy-decanoate). Although in these experiments some monomers, such as branched-chain, odd-chain or hydroxylated 3-hydroxyacids, were found included in PHAs after exogenous fatty acids were supplied to the transgenic plants, the same range of monomers would also be included in plant PHA from fatty acids supplied from endogenous fatty acid synthesis. Thus, one can predict being able to synthesize PHA polymers in plants that have a wide range of monomers, for example, higher proportion of short-chain monomers, unsaturated bonds at novel positions, monomers with hydroxylated groups, epoxy groups, acetylated groups, keto groups, cyclopentenyl groups, cyclopropanoid groups, furanoid groups or halogenated groups, branched chain, cyclic groups or any other novel monomers for which the equivalent functional groups exist in fatty acids in plants. The incorporation of these novel monomers derived from fatty acids into plant PHAs could be accomplished by expressing a PHA synthase in a plant which synthesizes these unusual fatty acids either naturally or after expression of a transgene such as fatty-acyl-thioesterases, -hydroxylases, -desaturases, -epoxidases, or -acetylases.

It is also conceivable that the substrate specificity of the PHA synthase could be modified to allow the incorporation of a wider range of monomers into PHA. One can predict that the range of monomers which could be included into plant PHAs from such a modified PHA synthase will include monomers that can be derived from plant fatty acid metabolism found in wild type plants or plants expressing transgenes (such as desaturases, hydroxylases, thioesterases, epoxidases, acetylases) which results in the modification of fatty acids synthesized in plants. It is also conceivable that suitable hydroxy acid substrates for the PHA synthase can be obtained from the amino acid metabolism or the plant secondary metabolism.

It has been demonstrated before that plants can synthesize PHB from acetyl-CoA through the expression of the 3-ketothiolase, acetoacetyl-CoA reductase and PHB synthase from *A. eutrophus* (Poitier, Y. et al., *Science* 256: 520-523 (1992); Nawrath, C. et al., *Proc. Natl. Acad. Sci. U.S.A.* 91: 12760-12764 (1994)). The examples herein demonstrate that PHA<sub>med</sub> can be synthesized in plants expressing a PHA synthase which can accept monomers from H6-H16. Since acetyl-CoA is also found in the peroxisome, one can predict that co-expression of a PHA synthase with a substrate specificity for 3-hydroxyacids ranging from H4 to H8 or higher in the peroxisome, and of the *A. eutrophus* acetoacetyl-CoA reductase, would lead to the biosynthesis of a copolymer containing hydroxybutyrate and hydroxyacids of H6 and higher. In this pathway, the expression of the 3-ketothiolase and hydroxyacids of H6 and higher. In this pathway, the expression of the 3-ketothiolase from *A. eutrophus* may not be required since the peroxisome already contains a 3-ketothiolase.

The examples herein clearly show that synthesis of PHA in plants can be significantly enhanced by increasing the pool of fatty acids which is channeled through  $\beta$ -oxidation. Thus, when short-chain fatty acids were added externally in the form of TWEEN-20 to PHA<sub>C1</sub>-transgenic plants, there was a 30- fold increase in the amount of PHA synthesized in plants. Similar large increases in PHA synthesis were found when tridecanoic acid and 8-methyl-nonanoic acid were added to the growth media. It is hypothesized that because these fatty acids could not be incorporated into membranes without disrupting them, the fatty acids are detoxified by channeling them to the peroxisome



for degradation by the  $\beta$ -oxidation cycle. Thus, increased channeling of fatty acids to the  $\beta$ -oxidation cycle results in an increase in PHA synthesized using intermediates of fatty acid oxidation. One can predict from this work that any changes in plants which results in an increased flux of fatty acids to the  $\beta$ -oxidation cycle will result in an increase in PHA synthesis in plants expressing a PHA synthase targeted to the peroxisome. Increasing the flux of fatty acids to the  $\beta$ -oxidation cycle could be accomplished by overexpressing enzymes which lead to the biosynthesis of modified fatty acids. This has been demonstrated in plants expressing thioesterase (Eccleston, V.S. et al., *Planta* 198: 46-53 (1996)) and implied in plants expressing hydroxylase (van de Loo, F.N. et al., *Proc. Natl. Acad. Sci. U.S.A.* 92: 6743-6747 (1995)). Increase of flux of lipids to the  $\beta$ -oxidation cycle and to PHA synthesis could also be accomplished by expressing other fatty acid modifying enzymes, such as desaturases, epoxidases, acetylases, enzymes involved in synthesis of branched-chain fatty acids, etcetera. This concept has been directly demonstrated in this present work with a fatty acyl-ACP thioesterase. It was shown that co-expression of a fatty acyl-ACP thioesterase in a plant expressing a peroxisomal PHA synthase leads to a 10 fold increase in PHA (Table 7). In addition of increasing the amount of PHA in plants, expression of the thioesterase leads to a predictable change in the composition of the PHA, i.e. since the *C. lanceolata* FatB3 thioesterase has the highest affinity for saturated C10 fatty acyl-ACP, there is a corresponding large increase in hydroxydecanoic acid (H10) present in the plant PHA (Table 8). Thus, expression of fatty acid modifying enzymes in conjunction with a PHA synthase in plants not only leads to an increase in the amount of PHA synthesized in plants, but also leads to a predictable changes in the PHA monomer composition, e.g. co-expression of a short-chain fatty acyl-ACP thioesterase would lead to an increase in the proportion of short-chain hydroxyacid monomers in plant PHA, co-expression of a long-chain fatty acyl-ACP thioesterase would lead to an increase in the proportion of long-chain hydroxyacid monomers in plant PHA, co-expression of a fatty acyl hydroxylase would lead to an increase in the proportion of hydroxylated hydroxyacid monomers in plant PHA, co-expression of a fatty acyl epoxidase would lead to an increase in the proportion of epoxidated monomers in plant PHA, co-expression of a fatty acyl acetylase would lead to an increase in the proportion of acetylated hydroxyacid monomers in plant PHA, and co-expression of a fatty acyl desaturase would lead to an increase in the

proportion of unsaturated hydroxyacid monomers in plant PHA. Increase in flux of lipids through the  $\beta$ -oxidation cycle could also be accomplished by overexpressing the key regulators (i.e. transcriptional factors) involved in the up-regulation of the entire  $\beta$ -oxidation cycle pathway during germination or senescence. This last approach would have the advantage of turning-on the  $\beta$ -oxidation cycle in tissues which normally have only low activity, such as the developing seeds of oil crops.

The examples herein point out the impact of fatty acid modifying enzymes for the production of novel PHA in transgenic plants expressing a PHA synthase. One key enzyme appears to be a 3-hydroxy-acyl-CoA epimerase. Although the normal function of the epimerase is to convert D-3-hydroxy-acyl-CoAs to the L-form required for the action of the L-3-hydroxy-acyl-CoA dehydrogenase, the reverse reaction of the epimerase can be responsible for converting the L-form to the D-form, which is essential for the activity of the PHA synthase. For that purpose the epimerase is important for the supply of the substrates for the PHA synthase derived from  $\beta$ -oxidation in the peroxisomes. Recombinant forms of such an epimerase activity expressed in peroxisomes or in other plant cell compartments like the cytoplasm or the plastids could play an important role in the production of PHA in transgenic plants. It is possible that the slow rate of the epimerase "reverse reaction" could be the major factor limiting the supply of substrates for the PHA synthase. The substrate limitation due to this could be the reason why PHA synthesis seemed to have reached a maximum in seedlings germinated both in the light and in the dark in liquid medium supplemented with TWEEN-20, which contains only saturated fatty acids.

The importance of certain fatty acid desaturases is highlighted by Table 3, wherein petroselinic acid (C18:1, 6-cis) was supplied to germinating PHAC1#3.3 seedlings in liquid medium, resulting in the specific increase of the H14 monomer. This indicated that any fatty acid containing unsaturated bonds starting at even-numbered carbons directly gives rise to the appropriate D-3-hydroxy-acyl-CoAs during  $\beta$ -oxidation, thus bypassing the otherwise necessary "reverse reaction" of the epimerase to generate the D-intermediates. Similarly the H8 and the H8:1 monomer are predicted to originate from the unsaturated fatty

acids linoleic acid (C18:2, 9,12-all cis) and linolenic acid (C18:3, 9,12,15-all cis). For that reason any plant containing high levels of fatty acids with unsaturated bonds starting at even-numbered carbons could be of interest for the production of PHA<sub>nm</sub>, or the transgenic expression of suitable fatty acid desaturases producing such unsaturated fatty acids in plants containing the PHA synthase would be similarly attractive for PHA production and monomer manipulation.

The examples herein demonstrate that a peroxisomally-located PHA synthase is able to divert intermediates from  $\beta$ -oxidation for their incorporation into PHA. The existence of the required D-3-hydroxy-acyl-CoA substrates was important for the synthesis of PHA. In light of the present disclosure, one may predict that PHA can be produced in a similar manner in any other compartment of any plant cell, provided that a supply of such D-3-hydroxy-acyl-CoA intermediates is present due either to an endogenous metabolic pathway or due to an artificially created pathway utilizing expression of transgenes. Fatty acid biosynthesis occurs in the plastids in plant cells, and modifications of this pathway could turn the plastids into a suitable source of D-3-hydroxy-acyl-CoA intermediates, which could subsequently be used to produce PHA either in the plastid itself or in other cell compartments.

#### EXAMPLE 17: Protein analysis

Leaves from transgenic plants were homogenized in 200 mM Tris-HCl (pH 7.5), 250 mM EDTA, 5 mM dithiothreitol and 1 mM phenylmethylsulfonyl fluoride. The homogenate was clarified by centrifugation and protein analyzed by Western blot using the ECL detection system (Amersham, Arlington Heights, IL).

#### EXAMPLE 18: Immunolocalization

Transgenic plants were grown on media containing MS salts, 1% sucrose, 0.7% agar and 50  $\mu$ g/mL kanamycin for either 7 days in the light or 1 day in the light followed by 6 days in the dark. Whole plants were fixed for 2 hours at room temperature in 4%

formaldehyde, 0.5% glutaraldehyde, 50 mM sodium cacodylate pH 7.3. The tissue samples were dehydrated in an ethanol series and embedded in LR White resin. Ultra thin sections were cut using a microtome, mounted on formvar-coated gold grids and blocked in 0.8% (w/v) bovine serum albumin, 0.1% (w/v) gelatine, 5% (w/v) normal goat serum and 2 mM sodium azide in PBS (10 mM sodium phosphate, 150 mM sodium chloride, pH 7.4). Grids were incubated for 1 hour at room temperature with antiserum against PHA synthase (1:50), glycolate oxidase (1:2000) and isocitrate lyase (1:1000) in the blocking solution followed by a 4 hour incubation at room temperature with a 1:50 dilution of gold-conjugated goat anti-rabbit antibodies (15 nm gold particles) in PBS. Immunolabeled sections were double-stained with uranyl acetate and lead citrate and viewed with a Jeol JEM transmission electron microscope.

#### EXAMPLE 19: PHA extraction and analysis

Fresh or dried frozen plant material was ground in a mortar and hyophlized. The powder was extracted with methanol in a Soxhlet apparatus for 24 hours followed by PHA extraction with chloroform for 24 hours, both at 85°C. The PHA-containing chloroform was concentrated under reduced pressure and extracted once with water to remove residual solid particles. PHA was precipitated by the addition of 10 volumes of cold methanol and subsequently washed by two cycles of chloroform solubilisation and methanol precipitation. PHA dissolved in chloroform was transesterified by acid methanolysis (Huijberts, G. N. et al., *Appl. Environ. Microbiol.* 58: 536-544 (1992)) and analyzed by gas-chromatography and mass spectrometry (GC-MS) using a Hewlett-Packard 5890 gas chromatograph (30 m long HP-5MS column) coupled to a Hewlett-Packard 5972 mass spectrometer (Hewlett Packard, Palo Alto, CA). Molecular weight determination of PHA samples were determined by gel permeation chromatography on a Waters 150 CV (Waters Corp., Milford, MA) equipped with a differential refractive index detector and an on-line viscometer and three ultrasyrigel columns in series ( $10^4$ ,  $10^5$  and  $10^6$  Å). Samples were prepared in dichloromethane and calibration performed using polystyrene standards.

EXAMPLE 20: Plant Vectors

In plants, transformation vectors capable of introducing encoding DNAs involved in PHA biosynthesis are easily designed, and generally contain one or more DNA coding sequences of interest under the transcriptional control of 5' and 3' regulatory sequences.

Such vectors generally comprise, operatively linked in sequence in the 5' to 3' direction, a promoter sequence that directs the transcription of a downstream heterologous structural DNA in a plant; optionally, a 5' non-translated leader sequence; a nucleotide sequence that encodes a protein of interest; and a 3' non-translated region that encodes a polyadenylation signal which functions in plant cells to cause the termination of transcription and the addition of polyadenylate nucleotides to the 3' end of the mRNA encoding said protein. Plant transformation vectors also generally contain a selectable marker. Typical 5'-3' regulatory sequences include a transcription initiation start site, a ribosome binding site, an RNA processing signal, a transcription termination site, and/or a polyadenylation signal. Vectors for plant transformation have been reviewed in Rodriguez et al. (Vectors: A Survey of Molecular Cloning Vectors and Their Uses, Butterworths, Boston. (1988)), Gluck et al. (Methods in Plant Molecular Biology and Biotechnology, CRC Press, Boca Raton, Fla. (1993)), and Croy (Plant Molecular Biology Labfax, Hanes and Rickwood (Eds.), BIOS Scientific Publishers Limited, Oxford, UK. (1993)).

EXAMPLE 21: Plant Promoters

Plant promoter sequences can be constitutive or inducible, environmentally- or developmentally-regulated, or cell- or tissue-specific. Often-used constitutive promoters include the CaMV 35S promoter (Odell et al., *Nature* 313: 810 (1985)), the enhanced CaMV 35S promoter, the Figwort Mosaic Virus (FMV) promoter (Richins et al., *Nucleic Acids Res.* 20: 8451 (1987)), the mannopine synthase (*mas*) promoter, the nopaline synthase (*nos*) promoter, and the octopine synthase (*ocs*) promoter. Useful inducible promoters include promoters induced by salicylic acid or polyacrylic acids (PR-1, Williams, S. W. et al, *Biotechnology* 10: 540-543 (1992)), induced by application of safeners (substituted benzenesulfonamide herbicides, Hershey, H.P. and Stoner, T.D., *Plant Mol. Biol.* 17: 679-

690 (1991)), heat-shock promoters (Ou-Lee et al., *Proc. Natl. Acad. Sci. U.S.A.* 83: 6815 (1986); Ainley et al., *Plant Mol. Biol.* 14: 949 (1990)), a nitrate-inducible promoter derived from the spinach nitrite reductase gene (Back et al., *Plant Mol. Biol.* 17: 9 (1991)), hormone-inducible promoters (Yamaguchi-Shinozaki et al., *Plant Mol. Biol.* 15: 905 (1990); Kares et al., *Plant Mol. Biol.* 15: 905 (1990)), and light-inducible promoters associated with the small subunit of RuBP carboxylase and LHCP gene families (Kuhlemeier et al., *Plant Cell* 1: 471 (1989); Feinbaum et al., *Mol. Gen. Genet.* 226: 449 (1991); Weisshaar et al., *EMBO J.* 10: 1777 (1991); Lam and Chua, *J. Biol. Chem.* 266: 17131 (1990); Castresana et al., *EMBO J.* 7: 1929 (1988); Schulze-Lefert et al., *EMBO J.* 8: 651 (1989)). Examples of useful tissue-specific, developmentally-regulated promoters include the  $\beta$ -conglycinin 7S promoter (Doyle et al., *J. Biol. Chem.* 261: 9228 (1986); Slighton and Beachy, *Plant* 172: 356 (1987)), and seed-specific promoters (Knutzon et al., *Proc. Natl. Acad. Sci. U.S.A.* 89: 2624 (1992); Bustos et al., *EMBO J.* 10: 1469 (1991); Lam and Chua, *Science* 248: 471 (1991); Stayton et al., *Aust. J. Plant. Physiol.* 18: 507 (1991)). Plant functional promoters useful for preferential expression in seed plastids include those from plant storage protein genes and from genes involved in fatty acid biosynthesis in oilseeds. Examples of such promoters include the 5' regulatory regions from such genes as napin (Kridl et al., *Seed Sci. Res.* 1: 209 (1991)), phaseolin, zein, soybean trypsin inhibitor, ACP, stearyl-ACP desaturase, and oleosin. Seed-specific gene regulation is discussed in EP 0 255 378. Promoter hybrids can also be constructed to enhance transcriptional activity (Comai, L. and Moran, P.M., U.S. Patent No. 5,106,739, issued April 21, 1992), or to combine desired transcriptional activity and tissue specificity.

EXAMPLE 22: Plant transformation and regeneration

A variety of different methods can be employed to introduce such vectors into plant protoplasts, cells, callus tissue, leaf discs, meristems, etcetera, to generate transgenic plants, including *Agrobacterium*-mediated transformation, particle gun delivery, microinjection, electroporation, polyethylene glycol-mediated protoplast transformation, liposome-mediated transformation, etc. (reviewed in Potrykus, *Ann. Rev. Plant Physiol. Plant Mol. Biol.* 42: 205 (1991)). In general, transgenic plants comprising cells containing and expressing

DNAs encoding enzymes facilitating PHA biosynthesis can be produced by transforming plant cells with a DNA construct as described above via any of the foregoing methods; selecting plant cells that have been transformed on a selective medium; regenerating plant cells that have been transformed to produce differentiated plants; and selecting a transformed plant which expresses the enzyme-encoding nucleotide sequence.

Specific methods for transforming a wide variety of dicots and obtaining transgenic plants are well documented in the literature (Gasser and Fraley, *Science* 244, 1293 (1989); Fisk and Dandekar, *Scientia Horticulturae* 55: 5 (1993); Christou, *Agro Food Industry Hi Tech*, p.17 (1994); and the references cited therein).

Successful transformation and plant regeneration have been reported in the monocots as follows: asparagus (*Asparagus officinalis*; Bytchier et al., *Proc. Natl. Acad. Sci. U.S.A.* 84: 5345 (1987)); barley (*Hordeum vulgare*; Wan and Lemaux, *Plant Physiol.* 104: 37 (1994)); maize (*Zea mays*; Rhodes et al., *Science* 240: 204 (1988); Gordon-Kamm et al., *Plant Cell* 2: 603 (1990); Fromm et al., *BioTechnology* 8: 833 (1990); Kozziel et al., *BioTechnology* 11: 194 (1993)); oats (*Avena sativa*; Somers et al., *BioTechnology* 10: 1589 (1992)); orchardgrass (*Dactylis glomerata*; Horn et al., *Plant Cell Rep.* 7: 469 (1988)); rice (*Oryza sativa*, including indica and japonica varieties; Toriyama et al., *BioTechnology* 6: 10 (1988); Zhang et al., *Plant Cell Rep.* 7: 379 (1988); Luo and Wu, *Plant Mol. Biol. Rep.* 6: 165 (1988); Zhang and Wu, *Theor. Appl. Genet.* 76: 835 (1988); Christou et al., *BioTechnology* 9: 957 (1991)); rye (*Secale cereale*; De la Pena et al., *Nature* 325: 274 (1987)); sorghum (*Sorghum bicolor*; Cassas et al., *Proc. Natl. Acad. Sci. USA* 90: 11212 (1993)); sugar cane (*Saccharum* spp.; Bower and Birch, *Plant J.* 2: 409 (1992)); tall fescue (*Festuca arundinacea*; Wang et al., *BioTechnology* 10: 691 (1992)); turfgrass (*Agrostis palustris*; Zhong et al., *Plant Cell Rep.* 13: 1 (1993)); wheat (*Triticum aestivum*; Vasil et al., *BioTechnology* 10: 667 (1992); Weeks et al., *Plant Physiol.* 102: 1077 (1993); Becker et al., *Plant J.* 5: 299 (1994)); and alfalfa (Masoud, S.A. et al., *Transgen. Res.* 5: 313 (1996)).

#### EXAMPLE 23: Host plants

Particularly useful plants for PHA production include those that produce carbon substrates which can be employed for PHA biosynthesis, including tobacco, wheat, potato, *Arabidopsis*, and high oil seed plants such as corn, soybean, canola, oil seed rape, sunflower, flax, peanut, sugarcane, switchgrass, and alfalfa.

If the host plant of choice does not produce the requisite fatty acid substrates in sufficient quantities, it can be modified, for example by mutagenesis or genetic transformation, to block or modulate the glycerol ester and fatty acid biosynthesis or degradation pathways so that it accumulates the appropriate substrates for PHA production. Expression of enzymes such as acyl-ACP thioesterase, fatty acyl hydroxylase, and yeast multifunctional protein (MFP) may serve to increase the flux of substrates in the peroxisome, leading to higher levels of PHA biosynthesis.

#### EXAMPLE 24: Nucleic acid mutation and hybridization

Variations in the nucleic acid sequence encoding a fusion protein may lead to mutant protein sequences that display equivalent or superior enzymatic characteristics when compared to the sequences disclosed herein. This invention accordingly encompasses nucleic acid sequences which are similar to the sequences disclosed herein, protein sequences which are similar to the sequences disclosed herein, and the nucleic acid sequences that encode them. Mutations may include deletions, insertions, truncations, substitutions, fusions, and the like.

Mutations to a nucleic acid sequence may be introduced in either a specific or random manner, both of which are well known to those of skill in the art of molecular biology. A myriad of site-directed mutagenesis techniques exist, typically using oligonucleotides to introduce mutations at specific locations in a nucleic acid sequence. Examples include single strand rescue (Kunkel, T. *Proc. Natl. Acad. Sci. U.S.A.*, 82: 488-492 (1985)), unique site elimination (Deng and Nickloff, *Anal. Biochem.* 200: 81 (1992)),

nick protection (Vandeyar, et al. *Gene* 65: 129-133 (1988)), and PCR (Costa, et al. *Methods Mol. Biol.* 57: 31-44 (1996)). Random or non-specific mutations may be generated by chemical agents (for a general review, see Singer and Kusmierek, *Ann. Rev. Biochem.* 52: 655-693 (1982)) such as nitrosoguanidine (Cerdeira-Olmedo et al., *J. Mol. Biol.* 33:705-719 (1968); Guerola, et al. *Nature New Biol.* 230: 122-125 (1971)) and 2-aminopurine (Rogan and Bessman, *J. Bacteriol.* 103: 622-633 (1970)), or by biological methods such as passage through mutator strains (Greener et al. *Mol. Biotechnol.* 7: 189-195 (1997)).

Nucleic acid hybridization is a technique well known to those of skill in the art of DNA manipulation. The hybridization properties of a given pair of nucleic acids is an indication of their similarity or identity. Mutated nucleic acid sequences may be selected for their similarity to the disclosed nucleic acid sequences on the basis of their hybridization to the disclosed sequences. Low stringency conditions may be used to select sequences with multiple mutations. One may wish to employ conditions such as about 0.15 M to about 0.9 M sodium chloride, at temperatures ranging from about 20°C to about 55°C. High stringency conditions may be used to select for nucleic acid sequences with higher degrees of identity to the disclosed sequences. Conditions employed may include about 0.02 M to about 0.15 M sodium chloride, about 0.5% to about 5% casein, about 0.02% SDS and/or about 0.1% N-laurylsarcosine, about 0.001 M to about 0.03 M sodium citrate, at temperatures between about 50°C and about 70°C. More preferably, high stringency conditions are 0.02 M sodium chloride, 0.5% casein, 0.02% SDS, 0.001 M sodium citrate, at a temperature of 50°C.

EXAMPLE 25: Determination of homologous and degenerate nucleic acid sequences

Modification and changes may be made in the sequence of the proteins of the present invention and the nucleic acid segments which encode them and still obtain a functional molecule that encodes a protein with desirable properties. The following is a discussion based upon changing the amino acid sequence of a protein to create an equivalent, or possibly an improved, second-generation molecule. The amino acid changes

may be achieved by changing the codons of the nucleic acid sequence, according to the codons given in Table 11.

Table 11: Codon degeneracies of amino acids

Amino acid	One letter	Three letter	Codons
Alanine	A	Ala	GCA GCC GCG GCT
Cysteine	C	Cys	TGC TGT
Aspartic acid	D	Asp	GAC GAT
Glutamic acid	E	Glu	GAA GAG
Phenylalanine	F	Phe	TTC TTT
Glycine	G	Gly	GGA GGC GGG GGT
Histidine	H	His	CAC CAT
Isoleucine	I	Ile	ATA ATC ATT
Lysine	K	Lys	AAA AAG
Leucine	L	Leu	TTA TTG CTA CTC CTG CTT
Methionine	M	Met	ATG
Asparagine	N	Asn	AAC AAT
Proline	P	Pro	CCA CCC CCG CCT
Glutamine	Q	Gln	CAA CAG
Arginine	R	Arg	AGA AGG CGA CGC CGG CGT
Serine	S	Ser	AGC AGT TCA TCC TCG TCT
Threonine	T	Thr	ACA ACC ACG ACT
Valine	V	Val	GTA GTC GTG GTT
Tryptophan	W	Trp	TGG
Tyrosine	Y	Tyr	TAC TAT

Certain amino acids may be substituted for other amino acids in a protein sequence without appreciable loss of enzymatic activity. It is thus contemplated that various changes may be made in the peptide sequences of the disclosed protein sequences, or their corresponding nucleic acid sequences without appreciable loss of the biological activity.

In making such changes, the hydropathic index of amino acids may be considered. The importance of the hydropathic amino acid index in conferring interactive biological function on a protein is generally understood in the art (Kyte and Doolittle, *J. Mol. Biol.* 157: 105-132 (1982)). It is accepted that the relative hydropathic character of the amino acid contributes to the secondary structure of the resultant protein, which in turn defines the

interaction of the protein with other molecules, for example, enzymes, substrates, receptors, DNA, antibodies, antigens, and the like.

Each amino acid has been assigned a hydrophobic index on the basis of their hydrophobicity and charge characteristics. These are: isoleucine (+4.5); valine (+4.2); leucine (+3.8); phenylalanine (+2.8); cysteine/cystine (+2.5); methionine (+1.9); alanine (+1.8); glycine (-0.4); threonine (-0.8); tryptophan (-0.9); tyrosine (-1.3); proline (-1.6); histidine (-3.2); glutamate/glutamine/aspartate/asparagine (-3.5); lysine (-3.9); and arginine (-4.5).

It is known in the art that certain amino acids may be substituted by other amino acids having a similar hydrophobic index or score and still result in a protein with similar biological activity, i.e., still obtain a biologically functional protein. In making such changes, the substitution of amino acids whose hydrophobic indices are within  $\pm 2$  is preferred, those within  $\pm 1$  are more preferred, and those within  $\pm 0.5$  are most preferred.

It is also understood in the art that the substitution of like amino acids may be made effectively on the basis of hydrophobicity. U.S. Patent No. 4,554,101 (Hopp, T.P., issued November 19, 1985) states that the greatest local average hydrophobicity of a protein, as governed by the hydrophobicity of its adjacent amino acids, correlates with a biological property of the protein. The following hydrophobicity values have been assigned to amino acids: arginine/lysine (+3.0); aspartate/glutamate (+3.0  $\pm 1$ ); serine (+0.3); asparagine/glutamine (+0.2); glycine (0); threonine (-0.4); proline (-0.5  $\pm 1$ ); alanine/histidine (-0.5); cysteine (-1.0); methionine (-1.3); valine (-1.5); leucine/isoleucine (-1.8); tyrosine (-2.3); phenylalanine (-2.5); and tryptophan (-3.4).

It is understood that an amino acid may be substituted by another amino acid having a similar hydrophobicity score and still result in a protein with similar biological activity, i.e., still obtain a biologically functional protein. In making such changes, the substitution of amino acids whose hydrophobic indices are within  $\pm 2$  is preferred, those within  $\pm 1$  are more preferred, and those within  $\pm 0.5$  are most preferred.

As outlined above, amino acid substitutions are therefore based on the relative similarity of the amino acid side-chain substituents, for example, their hydrophobicity, hydrophobicity, charge, size, and the like. Exemplary substitutions which take various of the foregoing characteristics into consideration are well known to those of skill in the art and include: arginine and lysine; glutamate and aspartate; serine and threonine; glutamine and asparagine; and valine, leucine, and isoleucine. Changes which are not expected to be advantageous may also be used if these resulted in functional fusion proteins.

All of the compositions and/or methods disclosed and claimed herein can be made and executed without undue experimentation in light of the present disclosure. While the compositions and methods of this invention have been described in terms of preferred embodiments, it will be apparent to those of skill in the art that variations may be applied to the compositions and/or methods and in the steps or in the sequence of steps of the methods described herein without departing from the concept, spirit and scope of the invention. More specifically, it will be apparent that certain agents which are both chemically and physiologically related may be substituted for the agents described herein while the same or similar results would be achieved. All such similar substitutes and modifications apparent to those skilled in the art are deemed to be within the spirit, scope and concept of the invention.

## SEQUENCE LISTING

## SEQUENCE LISTING

## (1) GENERAL INFORMATION:

## (i) APPLICANT:

(A) NAME: VOLKER MITTENDORF  
 (B) STREET: Institut de Biologie et Physiologie Vegetales  
 (C) CITY: Batiment de Biologie  
 (D) STATE: Lausanne  
 (E) COUNTRY: Switzerland  
 (F) POSTAL CODE (ZIP): CH-1015  
 (G) TELEPHONE: (41) (21) 692-4222  
 (H) TELEFAX: (41) (21) 692-4195

(A) NAME: YVES POIRIER  
 (B) STREET: Institut de Biologie et Physiologie Vegetales  
 (C) CITY: Batiment de Biologie  
 (D) STATE: Lausanne  
 (E) COUNTRY: Switzerland  
 (F) POSTAL CODE (ZIP): CH-1015  
 (G) TELEPHONE: (41) (21) 692-4222  
 (H) TELEFAX: (41) (21) 692-4195

## (ii) TITLE OF INVENTION: BIOSYNTHESIS OF MEDIUM CHAIN LENGTH POLYHYDROXYALKANATES

## (iii) NUMBER OF SEQUENCES: 26

## (iv) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk  
 (B) COMPUTER: IBM PC compatible  
 (C) OPERATING SYSTEM: PC-DOS/MS-DOS  
 (D) SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

## (2) INFORMATION FOR SEQ ID NO: 1:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1677 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

ATGAGTCAGA AGACAAATRA CGAGCTTCCC AGGCAGCCG CGGAATACAC GCTGAACCTG 60  
 AATCCGCTGA TCGCATCCG GGCAGAGAC CTGCTACCT CCGCGCGCAT GGTCTGCTC 120  
 CAGCGGTGC GCGAGCGCT GCACAGGCC AGGCACGTGG CGCATTTGAG CTTGGAGCTG 180

-59-

AAGAAGCTCC TGCTCGCCA GTCCGAGCTA CGCCAGGCG ATGACGACCG AGGCTTTTCC 240  
 GATCGGCTT GGAGCCAGAA TCCACTGTATC AAGCGTACA TGCAGACCTA CTTGGCCTGG 300  
 CCGAAGGAGC TGCACAGCTG GATCAGCCAC AGGACCTGT CGCCGAGGA CATCAGTGT 360  
 GGCAGTTTCG TCATCAACCT GCTGACCGAG GCGATGTGC CGACCAACAG CTTGAGCAAC 420  
 CCGCGGCGG TCAAGCGCTT CTTGAGACC GCGCGCAGA GCTGTCTGGA CGGCTCTGGC 480  
 CACTTGGCCA AGGACCTGTG GAACAAGGC GGGATCCGA GCCAGGTGGA CATGAGGCC 540  
 TTGAGGTGG GCAAGAACTT GGCACACCC GAGGCGCGG TGGTGTTCG CAACGAGCTG 600  
 CTGGAACCTGA TCCAGTACCG GCCGATCAC GAGTCTGTG ACGAACGCC GCTGCTGTG 660  
 GTCCGCGCG AGATCAACAA GTTCTAGTC TTGACCTGT CGCCGACAA GAGCCTGGCG 720  
 CGCTTCTGCC TGCAGAACG CGTGCAGACC TTCATGCTA GTTGGGCAA CCGGACCAAG 780  
 TGCAGCGCG AATGGGCTT GACCACCTAT ATCAGGCGC TCAAGGAGC CATCGAGTA 840  
 GTCTGTGCA TCACCGGCG CAGGACCTC AACCTCTCG GCGCTGTCT CCGCGGATC 900  
 ACCACCGGA CCTGTGTCG CCACTAGTG GCACGCGCG AGAAGAGGT CAACGCTTC 960  
 ACCCACTGG TCAGGTGCT CCACTTGGAA CTGAATACC AGGTGCGCT GTTCGCGAC 1020  
 GAGAAGACTC TGAAGGCGC CAGCGTGT TCTTACCAT CCGCGTGTCT GGAGGCAAG 1080  
 GACATGGCA AGGTGTTGCG CTGGATGCG CCCACGACC TGATCTGGA CTACTGGGTC 1140  
 AACAACTACC TGTCTGGCAA CAGCGCGCG CGTTTGACA TCTCTTACTG GAACACGAC 1200  
 ACCACGCGC TGCCGCGCG GCTGCAGCG GAGTTGTCG AACTGTTCA GAGCAACCG 1260  
 CTGAACCGC CCGCGGCTT GGAGTCTCC GGCACGCCA TCGACCTGA GCAAGTACT 1320  
 TGCAGCTTCT ACTGTGTCG CGGTCTGAC GACCACATCA CCCCCGGA GTGCTGCTAC 1380  
 AAGTCGCGCA GGTGCTGGG TGGCAAGTGC GAGTTTATCC TCTCCAACAG CGGTACATC 1440  
 CAGAGATCC TCACCCACC GGGCAACCCC AAGGCACGCT TCATGACCA TCCGGAATG 1500  
 CCGCGCGAG CCAAGGCGT GCTGGAACAG CCGGCGACG ACCGCGACTC GTGGTGGTGG 1560  
 CACTGGCAG AATGCTGGC CGAAGCTTC GGCAGACCC GCAAGGCGC CGCCAGCTG 1620  
 GGCACAGA CTTATCCGC CCGCGAGCC GCGCCCGGA CTTACGTGCA TGAACGA 1677

## (2) INFORMATION FOR SEQ ID NO: 2:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 559 amino acids  
 (B) TYPE: amino acid

-60-

(C) STRANDEDNESS:  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

```

5 Met Ser Gln Lys Asn Asn Asn Gln Leu Pro Lys Gln Ala Ala Gln Asn
  1      5      10      15
10 Thr Leu Asn Leu Asn Pro Val Ile Gly Ile Arg Gly Lys Asp Leu Leu
  20      25      30
15 Thr Ser Ala Arg Met Val Leu Leu Gln Ala Val Arg Gln Pro Leu His
  35      40      45
15 Ser Ala Arg His Val Ala His Phe Ser Leu Gln Leu Lys Asn Val Leu
  50      55      60
65 Leu Gly Gln Ser Gln Leu Arg Pro Gly Asp Asp Arg Arg Phe Ser
  70      75      80
20 Asp Pro Ala Trp Ser Gln Asn Pro Leu Leu Tyr Lys Arg Tyr Met Gln Thr
  85      90      95
25 Tyr Leu Ala Trp Arg Lys Gln Leu His Ser Trp Ile Ser His Ser Asp
  100     105     110
25 Leu Ser Pro Gln Asp Ile Ser Arg Gly Gln Phe Val Ile Asn Leu Leu
  115     120     125
30 Thr Gln Ala Met Ser Pro Thr Asn Ser Leu Ser Asn Pro Ala Ala Val
  130     135     140
35 Lys Arg Phe Phe Gln Thr Gly Gly Lys Ser Leu Leu Asp Gly Leu Gly
  145     150     155     160
35 His Leu Ala Lys Asp Leu Val Asn Asn Gly Gly Met Pro Ser Gln Val
  165     170     175
35 Asp Met Asp Ala Phe Gln Val Gly Lys Asn Leu Ala Thr Thr Gln Gly
  180     185     190
40 Ala Val Val Phe Arg Asn Asp Val Leu Gln Leu Ile Gln Tyr Arg Pro
  195     200     205
45 Ile Thr Gln Ser Val His Gln Arg Pro Leu Leu Val Val Pro Pro Gln
  210     215     220
45 Ile Asn Lys Phe Tyr Val Phe Asp Leu Ser Pro Asp Lys Ser Leu Ala
  225     230     235     240
50 Arg Phe Cys Leu Arg Asn Gly Val Gln Thr Phe Ile Val Ser Trp Arg
  245     250     255
50 Asn Pro Thr Lys Ser Gln Arg Gln Trp Gly Leu Thr Thr Tyr Ile Gln
  260     265     270
55 Ala Leu Lys Gln Ala Ile Gln Val Val Leu Ser Ile Thr Gly Ser Lys

```

-61-

```

5 Asp Leu Asn Leu Leu Gly Ala Cys Ser Gly Gly Ile Thr Thr Ala Thr
  275      280      285
10 Leu Val Gly His Tyr Val Ala Ser Gly Gln Lys Lys Val Asn Ala Phe
  290      295      300
10 Thr Gln Leu Val Ser Val Leu Asp Phe Gln Leu Asn Thr Gln Val Ala
  305      310      315      320
10 Leu Phe Ala Asp Gln Lys Thr Leu Gln Ala Ala Lys Arg Arg Ser Tyr
  325      330      335
15 Gln Ser Gly Val Leu Gln Gly Lys Asp Met Ala Lys Val Phe Ala Trp
  340      345      350
15 Met Arg Pro Asn Asp Leu Ile Trp Asn Tyr Trp Val Asn Asn Tyr Leu
  355      360      365
20 Leu Gly Asn Gln Pro Pro Ala Phe Asp Ile Leu Tyr Trp Asn Asn Asp
  370      375      380
20 Thr Thr Arg Leu Pro Ala Ala Leu His Gly Gln Phe Val Gln Leu Phe
  385      390      395      400
25 Lys Ser Asn Pro Leu Asn Arg Pro Gly Ala Leu Gln Val Ser Gly Thr
  405      410      415
30 Pro Ile Asp Leu Lys Gln Val Thr Cys Asp Phe Tyr Cys Val Ala Gly
  420      425      430
30 Leu Asn Asp His Ile Thr Pro Trp Gln Ser Cys Tyr Lys Ser Ala Arg
  435      440      445
35 Leu Leu Gly Gly Lys Cys Gln Phe Ile Leu Ser Asn Ser Gly His Ile
  450      455      460
35 Gln Ser Ile Leu Asn Pro Pro Gly Asn Pro Lys Ala Arg Phe Met Thr
  465      470      475      480
40 Asn Pro Gln Leu Pro Ala Gln Pro Lys Ala Trp Leu Gln Gln Ala Gly
  485      490      495
45 Lys His Ala Asp Ser Trp Trp Leu His Trp Gln Gln Trp Leu Ala Gln
  500      505      510
45 Arg Ser Gly Lys Thr Arg Lys Ala Pro Ala Ser Leu Gly Asn Lys Thr
  515      520      525
50 Tyr Pro Ala Gly Gln Ala Ala Pro Gly Thr Tyr Val His Gln Arg
  530      535      540
55 Tyr Pro Ala Gly Gln Ala Ala Pro Gly Thr Tyr Val His Gln Arg
  545      550      555

```

(2) INFORMATION FOR SEQ ID NO: 3:

(1) SEQUENCE CHARACTERISTICS:

-62-



(A) LENGTH: 1680 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

TATCGTTCGG CCTCTCTGCT GGGCGGCCAG CCGCCCTTCA TCCTGTCCAA CAGCGGCAC 1440  
ATCCAGACGA TCCTCAACCC TCCCGGAAAC CCCAGAGCCT GCTACTTCGA GACACACAG 1500  
CTGAGCAGCG ATCCACGGCG CTGTACTTAC GACGCCAAGC GCGAAGAGGG CAGCTGGTGG 1560  
CCGGTCTGGC TGGGCTGGCT GCAGGAGCG TCGGGCGAGC TGGGCAACCC TGACTTCAAC 1620  
CTTGGCAGCG CCGGCATCC GCCCTTCGAA GGGGCCCGG GCACCTAGCT GCATATAGCG 1680

## (2) INFORMATION FOR SEQ ID NO: 4:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 560 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS:

(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

Met Arg Glu Lys Gln Glu Ser Gly Ser Val Pro Val Pro Ala Glu Phe 1 15  
Met Ser Ala Gln Ser Ala Ile Val Gly Leu Arg Gly Lys Asp Leu Leu 20 30  
Thr Thr Val Arg Ser Leu Ala Val His Gly Leu Arg Gln Pro Leu His 35 45  
Ser Ala Arg His Leu Val Ala Phe Gly Gly Gln Leu Gly Lys Val Leu 50 60  
Leu Gly Asp Thr Leu His Gln Pro Asn Pro Gln Asp Ala Arg Phe Gln 65 80  
Asp Pro Ser Trp Arg Leu Asn Pro Phe Tyr Arg Arg Thr Leu Gln Ala 85 95  
Tyr Leu Ala Trp Gln Lys Gln Leu Leu Ala Trp Ile Asp Glu Ser Asn 100 110  
Leu Asp Cys Asp Asp Arg Ala Arg Ala Arg Phe Leu Val Ala Leu Leu 115 125  
Ser Asp Ala Val Ala Pro Ser Asn Ser Leu Ile Asn Pro Leu Ala Leu 130 140  
Lys Glu Leu Phe Asn Thr Gly Gly Ile Ser Leu Leu Asn Gly Val Arg 145 155 160

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

ATCGAGAA AGCAGGAATC GGTAGCGTG CCGGTGCCCG CCGAGTTTCA GAGTGCACAG 60  
AGGCCATCG TCGGCTTCGG CCGCAAGGAC CTGTGACGA CGGTCCGCAG CCTGGCTGTC 120  
CAGGCGCTCG GCCAGCCGCT GCACAGTGG CCGCACTTGG TCGCTTTCGG AGGCCAGTTG 180  
GGCAGGTGC TGTGGGGCA CACCTGAC CAGCCGAACC CACAGAGCG CCGTTCCAG 240  
GATCCATCCT GCGGCCCTCA TCCTTCTTAC CCGCGCACCC TGCAGGCCTA CCTGGCGTGG 300  
CAGAAACAC TGTGCGCTG GATCGAGAA AGCAACCTGG ACTGACGA TCGCGCCCGC 360  
GCCGCTTCC TCGTGGCTT GCTCTCCGC GCCGTGGAC CCGCAACAG CTTGATCAAT 420  
CCACTGCGCT TAAAGAACT GTTCAATACC GCGGGATCA GCCTGTCTCA TGGCGTCCGC 480  
CACCTGCTG AAGACCTGCT GCACACGCG GGCATGCCCC GCCAGGTAA CAAGACCGCC 540  
TTGAGATCG GTGCAACT CGCCACACG CAGGCGCGG TGGTGTCCG CACGAGGTG 600  
CTGGAGTCA TCCAGTACAA GCGGCTGGC GAGGCCAGT AGGCCAAGCC CTTGCTGATC 660  
GTGCGCCGC AGATCAACAA GTACTACATC TTGACCTGT CCGCGGAAA GAGCTTCGTC 720  
CAGTACGCC TGAAGAACAA CTTGACAGTC TTGCTCATCA GTTGGCGAA CCGCGACGCC 780  
CAGCACCGG ATGGGGCT GAGCACTAT GTCGAGGCC TCGACCAAGC CATCGAGTTC 840  
AGCGCGAGA TCACCGGCAG CCGCAGCGT AACCTGGCG GCGCTGCGC CCGCGGCTC 900  
ACCGTAGCG CCTTGCTCG CCACCTGCG GTGCGCCGGC AACGCGCAA GGTCAGTAGC 960  
GTCACTTAC TGTGAGCT GCTGACAGC CAGATGAAA GCCCGGCGAT GCTCTTCGCC 1020  
GACGAGCAGA CCTTGAGAG CAGCAAGCG CGCTCTACC AGCATGGCT GCTGGACGG 1080  
CGCGACATGG CCAAGGTGT CGCTGATG CCGCCCAACG ACCTGATCTG GAATCTACTGG 1140  
GTCAACACT ACTGTGCTGG CAGGCGGCCG CCGGCTTTCG ACATCTCTTA CTGGAACAC 1200  
GACACACCG GGTGCGCCG GCGTTCCAC GCGCACTGC TCGACCTGTT CAAGCACAC 1260  
CCGCTGACCC GCCCGGCGC GCTGGAGTC AGCGGACCG CCGTGGACCT GGGCAAGGTG 1320  
GCGATGACA GTTCCAGT CCGCGGCATC ACCGACCACA TCAGGCCCTG GAGCGCGTG 1380

His Leu Leu Glu Asp Leu Val His Asn Gly Met Pro Ser Gln Val  
165 170 175  
Asn Lys Thr Ala Phe Glu Ile Gly Arg Asn Leu Ala Thr Thr Gln Gly  
180 185 190  
Ala Val Val Phe Arg Asn Glu Val Leu Glu Leu Ile Gln Tyr Lys Pro  
195 200 205  
Leu Gly Glu Arg Gln Tyr Ala Lys Pro Leu Leu Ile Val Pro Pro Gln  
210 215 220  
Ile Asn Lys Tyr Tyr Ile Phe Asp Leu Ser Pro Glu Lys Ser Phe Val  
225 230 235 240  
Gln Tyr Ala Leu Lys Asn Asn Leu Gln Val Phe Val Ile Ser Tyr Arg  
245 250 255  
Asn Pro Asp Ala Gln His Arg Glu Trp Gly Leu Ser Thr Tyr Val Glu  
260 265 270  
Ala Leu Asp Gln Ala Ile Glu Val Ser Arg Glu Ile Thr Gly Ser Arg  
275 280 285  
Ser Val Asn Leu Ala Gly Ala Cys Ala Gly Gly Leu Thr Val Ala Ala  
290 295 300  
Leu Leu Gly His Leu Gln Val Arg Arg Gln Leu Arg Lys Val Ser Ser  
305 310 315 320  
Val Thr Tyr Leu Val Ser Leu Leu Asp Ser Gln Met Glu Ser Pro Ala  
325 330 335  
Met Leu Phe Ala Asp Glu Gln Thr Leu Glu Ser Ser Lys Arg Arg Ser  
340 345 350  
Tyr Gln His Gly Val Leu Asp Gly Arg Asp Met Ala Lys Val Phe Ala  
355 360 365  
Trp Met Arg Pro Asn Asp Leu Ile Trp Asn Tyr Trp Val Asn Asn Tyr  
370 375 380  
Leu Leu Gly Arg Gln Pro Pro Ala Phe Asp Ile Leu Tyr Trp Asn Asn  
385 390 395 400  
Asp Asn Thr Arg Leu Pro Ala Ala Phe His Gly Glu Leu Leu Asp Leu  
405 410 415  
Phe Lys His Asn Pro Leu Thr Arg Pro Gly Ala Leu Glu Val Ser Gly  
420 425 430  
Thr Ala Val Asp Leu Gly Lys Val Ala Ile Asp Ser Phe His Val Ala  
435 440 445  
Gly Ile Thr Asp His Ile Thr Pro Trp Asp Ala Val Tyr Arg Ser Ala  
450 455 460

-65-

Leu Leu Leu Gly Gly Gln Arg Arg Phe Ile Leu Ser Asn Ser Gly His  
465 470 480  
Ile Gln Ser Ile Leu Asn Pro Pro Gly Asn Pro Lys Ala Cys Tyr Phe  
485 490 495  
Glu Asn Asp Lys Leu Ser Ser Asp Pro Arg Ala Trp Tyr Tyr Asp Ala  
500 505 510  
Lys Arg Glu Glu Gly Ser Trp Trp Pro Val Trp Leu Gly Trp Leu Gln  
515 520 525  
Glu Arg Ser Gly Glu Leu Gly Asn Pro Asp Phe Asn Leu Gly Ser Ala  
530 535 540  
Ala His Pro Pro Leu Glu Ala Ala Pro Gly Thr Tyr Val His Ile Arg  
545 550 555 560  
(2) INFORMATION FOR SEQ ID NO: 5:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 29 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear  
(2) INFORMATION FOR SEQ ID NO: 6:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 29 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear  
(2) INFORMATION FOR SEQ ID NO: 7:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 29 base pairs

WO 99/35278

PCT/US98/00083

(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

GTGGAATTCA TCGGTGAAAA GCAGGAATC

10 (2) INFORMATION FOR SEQ ID NO: 8:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 29 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

GGCCAAGCTT TTAGCGTAT ATGCACGTA

25 (2) INFORMATION FOR SEQ ID NO: 9:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1731 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

40 ATGGCTGCAT CTTTCTCTGT CCCCTCTATG ATCATGGAAG AGGAGGGAG ATTGAGGCG

GAAGTTGCGG AAGTGCAGAC TTGGTGGAGC TCAGAGAGGT TCAAGCTAAC AAGCGTCTCT

TACACGGCCC GTGACGTGGT GGCCTACGTT GGTCACTTCA AGCAGGTTTA TGCTTCBAAC

45 GAGATGGCTA AGAAGCTGTG GAGAACGCTC AAGAGTCACC AAGTCAACGG CACGGCGTCT

CGCAGCTTGG GTGCCTTGGG CCTGTTTCAG GTGACATAGA TGGCTAAACA TTTAGACACC

50 ATTATGCT CTGGTTGGCA GTGCTCGTCT ACTCACACTT CCATTAACGA GCCTGGTCCG

GATCTTGTCT ACTATCCATA CGATACCGTT CTTAACAAAG TCGAACATCT CTTCTTCGCT

CACAGTACC ATGACAGAAA ACAGAGGGAG GCGAGATAGA CATGAGCAG AGAGAGAGA

55 GCBAABACTC CGTTTGTGGA CTACTTGAG CCATCATCG CCGACGGAG AACCGGCTTC

-67-

WO 99/35278

PCT/US98/00083

GGCGGTACCA CTGCCACCTT AAACTCTTGC AAACTCTTGC TTGAAGAGG AGCGCTGGG

5 GTCCACATCG AGGACCAATC CTCCTCACC AAGAAGTGTG GCCACATGCC CGGAAAAGTC

CTCGTGGCAG TCAGTGACA CATTACACGG CTTGTTGGGG CTGGCTCCA GTTCAGCTG

ATGGGCACAG AGACCGTCTT GGTGCTAGA ACGGACGGG TCGCGCCAC TCTGATCCAA

10 TCGAACATTG ACTCAAGGA CCACCAATC ATCTCGGTG TCACCTAGC AAACCTTAGA

GCGAAGATT TGTCTCGCT TCTGGCCGAG GGAATGGCTG TAGGCAATTA TGGTCCAGCG

TTGCAAGCGA TTGAGGATCA ATGCTTAGC TCAGCTGCTC TCATGACTTT CTCGACGCT

15 GTCGTGGAGG CTCTCAAGCG CATGAACCTA AGTGAAGT AGAAGAGCG GAGATGACC

GAGTGGCTTA TCCATGCAAG GTACGAGAAC TGCCTTTCAA ACAGCAAGG CCGAGATTTA

20 GCAGCAAAAC TCGGTGTGAC TGATCTTTTC TGGGACTGGG ACTTGCCAG AACCAAGAA

GGATTCTACC GGTTCCAAGG CTGGGTACA GCAGCGGTG TCGGTGGCTG GGCCTTTGCA

CAGATAGCTG ATCTCATCTG GATGGAACG GCAAGCCCTG ACTCAACGA ATGCACCAA

25 TTGCGAAG GAGTCAAGTC CAGACACCA GAGTATGC TGGCTTCAA CTTCTCCCA

TCCTTCACT GGGACGCTTC TGATATGAG GATCAGAGA TGATGAGTT CATTCACGA

30 ATGCCAGGC TCGGTATTG CTGGCAGTTT ATAACTTTC CCGGTTTCCA TCGGATGCT

CTTGTTGTCG ATACGTTTGC AAGGATTAAC GCGAGGAGAG GATGCTGGC TTATGTCAG

AGGATACAGA GAGAGAGAG GAGCATGGG GTTGACACAT TGCTCATCA GAAATGGTCA

35 GGTGCTAATT ACTATGATCG TTATCTTAAG ACGTCCAG GTGGAATCT CTCCTACTGA

GCCATGGGCA AAGGTGTTAC CGAGGAACAA TTCAAAGAGA CTGAGACGAG GCGGGAGCT

40 GCTGGAATGG GCGAAGGAG TAGCTTGTG GTGGCCAGT CCAGATGTA A

(2) INFORMATION FOR SEQ ID NO: 10:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 576 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS:  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

Met Ala Ala Ser Phe Ser Val Pro Ser Met Ile Met Glu Glu Gly

-68-

1 5 10 15  
 Arg Phe Glu Ala Glu Val Ala Glu Val Gln Thr Trp Trp Ser Ser Glu  
 20 25 30  
 5 Arg Phe Lys Leu Thr Arg Arg Pro Tyr Thr Ala Arg Asp Val Val Ala  
 35 40 45  
 10 Leu Arg Gly His Leu Lys Gln Gly Tyr Ala Ser Asn Glu Met Ala Lys  
 50 55 60  
 65 Lys Leu Trp Arg Thr Leu Lys Ser His Gln Val Asn Gly Thr Ala Ser  
 70 75 80  
 15 Arg Thr Phe Gly Ala Leu Asp Pro Val Gln Val Thr Met Met Ala Lys  
 85 90 95  
 100 His Leu Asp Thr Ile Tyr Val Ser Gly Trp Gln Cys Ser Ser Thr His  
 105 110  
 20 Thr Ser Thr Asn Glu Pro Gly Pro Asp Leu Ala Asp Tyr Pro Tyr Asp  
 115 120 125  
 130 Thr Val Pro Asn Lys Val Glu His Leu Phe Phe Ala Gln Gln Tyr His  
 135 140  
 25 Asp Arg Lys Gln Arg Glu Ala Arg Met Ser Met Ser Arg Glu Glu Arg  
 145 150 155 160  
 30 Ala Lys Thr Pro Phe Val Asp Tyr Leu Lys Pro Ile Ile Ala Asp Gly  
 165 170 175  
 180 Gly Thr Gly Phe Gly Gly Thr Thr Ala Thr Val Lys Leu Cys Lys Leu  
 185 190  
 35 Phe Val Glu Arg Gly Ala Ala Gly Val His Ile Glu Asp Gln Ser Ser  
 195 200 205  
 40 Val Thr Lys Lys Cys Gly His Met Ala Gly Lys Val Leu Val Ala Val  
 210 215 220  
 225 Ser Glu His Ile Asn Arg Leu Val Ala Ala Arg Leu Gln Phe Asp Val  
 230 235 240  
 45 Met Gly Thr Glu Thr Val Leu Val Ala Arg Thr Asp Ala Val Ala Pro  
 245 250 255  
 260 Thr Leu Ile Gln Ser Asn Ile Asp Ser Arg Asp His Gln Phe Ile Leu  
 265 270  
 50 Gly Val Thr Asn Pro Asn Leu Arg Gly Lys Ser Leu Ser Ser Leu Leu  
 275 280 285  
 290 Ala Glu Gly Met Ala Val Gly Asn Asn Gly Pro Ala Leu Gln Ala Ile  
 295 300

-69-

Glu Asp Gln Trp Leu Ser Ser Ala Arg Leu Met Thr Phe Ser Asp Ala  
 305 310 315 320  
 5 Val Val Glu Ala Leu Lys Arg Met Asn Leu Ser Glu Asn Glu Lys Ser  
 325 330 335  
 Arg Arg Val Thr Glu Trp Leu Ile His Ala Arg Tyr Glu Asn Cys Leu  
 340 345 350  
 10 Ser Asn Glu Gln Gly Arg Glu Leu Ala Ala Lys Leu Gly Val Thr Asp  
 355 360 365  
 15 Leu Phe Trp Asp Trp Asp Leu Pro Arg Thr Arg Glu Gly Phe Tyr Arg  
 370 375 380  
 385 Phe Gln Gly Ser Val Thr Ala Ala Val Val Arg Gly Trp Ala Phe Ala  
 390 395 400  
 20 Gln Ile Ala Asp Leu Ile Trp Met Glu Thr Ala Ser Pro Asp Leu Asn  
 405 410 415  
 420 Glu Cys Thr Gln Phe Ala Glu Gly Val Lys Ser Lys Thr Pro Glu Val  
 425 430  
 25 Met Leu Ala Tyr Asn Leu Ser Pro Ser Phe Asn Trp Asp Ala Ser Gly  
 435 440 445  
 450 Met Thr Asp Gln Gln Met Met Glu Phe Ile Pro Arg Ile Ala Arg Leu  
 455 460  
 30 Gly Tyr Cys Trp Gln Phe Ile Thr Leu Ala Gly Phe His Ala Asp Ala  
 465 470 475 480  
 485 Leu Val Val Asp Thr Phe Ala Lys Asp Tyr Ala Arg Arg Gly Met Leu  
 490 495  
 35 Ala Tyr Val Glu Arg Ile Gln Arg Glu Glu Arg Ser Asn Gly Val Asp  
 500 505 510  
 40 Thr Leu Ala His Gln Lys Trp Ser Gly Ala Asn Tyr Tyr Asp Arg Tyr  
 515 520 525  
 530 Leu Lys Thr Val Gln Gly Gly Ile Ser Ser Thr Ala Ala Met Gly Lys  
 535 540  
 45 Gly Val Thr Glu Glu Gln Phe Lys Glu Thr Thr Arg Pro Gly Ala  
 545 550 555 560  
 565 Ala Gly Met Gly Glu Gly Thr Ser Leu Val Val Ala Lys Ser Arg Met  
 570 575

(2) INFORMATION FOR SEQ ID NO: 11:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 26 base pairs

-70-

- (B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

5  
10  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:  
ACTGAGCTT TGGCAAAGG TGTAC 26

- (2) INFORMATION FOR SEQ ID NO: 12:  
(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 28 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

25  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:  
GTGGTCTAGA AGTTTTCTG CGAAGATG 28

- 30  
(2) INFORMATION FOR SEQ ID NO: 13:  
(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 102 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

35  
40  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:  
GGCAAGGTG TTACCGAGGA ACATTCAAA GAGACTGGA CGAGGCCGGG AGCTGCTGGA 60

45  
102  
ATGGGCGAAG GGAAGTACCT TGTGTGGCC AAGTCCAGAA TG

- (2) INFORMATION FOR SEQ ID NO: 14:  
(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 34 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS:  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

5  
10  
15  
Gly Lys Gly Val Thr Glu Gln Phe Lys Glu Thr Trp Thr Arg Pro  
1 5 10 15  
20  
25  
30  
Gly Ala Ala Gly Met Gly Glu Thr Ser Leu Val Val Ala Lys Ser  
10 Arg Met 26

- (2) INFORMATION FOR SEQ ID NO: 15:  
(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1677 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

25  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

ATGAGCCAGA AGAACAATTA CAGCTTCCC AAGCAAGCG CGGAAACAC GGTGAACCTG 60  
AATCCGTGA TCGGCATCCG GGGCAGGAC CTGCTCACT CGCGCGCAT GTTCTGTCTC 120  
CAGGCGGTGC GCGAGCGCT GCACAGCGC AGGACGTGG CGCATTTTCA CCTGGAGCTG 180  
AAGAACGTCC TGCTCGGCCA GTCGAGCTA CGCCAGGCG ATGACGACCG ACGCTTTTCC 240  
GATCCGGCTT GGAGCCAGAA TCCACTGTAC AAGGCTTACA TGCAGACCTA CTTGGCCTGG 300  
CGCAAGGAGC TGCACAGCTG GATCAGCCAC AGCGACTGT CGCCGAGGA CATCAGTCGT 360  
GGCCAGTTGG TCATCAACTT GCTGACCGAG GCGATGTGCG CGACCAACAG CTTGAGCAAC 420  
CCGGCGGCGG TCAAGCGCTT CTTGAGACC GCGGCGAAGA GCTGTCTGGA CGGCTCGGC 480  
CACTTGGCCA AGGACCTGTT GAACAGGCG GGGATGCCGA GCCAGGTGGA CATGACGCC 540  
TTGAGAGTGG GCAAGAACCT GCCCACCACC GAGGCGCGCG TGTGTGTTCCG CAACGACGTG 600  
CTGGAATGTA TCCAGTACCG GCCGATACC GAGTCGGTGC ACGAACGCC GCTGTGTTGTG 660  
GTGCGCGCGC AGATCAACAA GTTCTAGCTC TTGACCTGTT CGCCGGACAA GAGCTTGGCG 720  
CGCTTCTGCC TCGCAACGCG CGTGACAGCC TTCAATGTCA GTTGGCGCA CCGACCAAG 780  
TGCAGCGCG AATGGGCTT GACCACCTAT ATCGAGGCGC TCAAGGAGGC CATCGAGGTA 840  
GTCCTGTGTA TCACCGGAG CAGGACCTC AACCTCTCG GCGCTGTCTC CGGCGGGATC 900

ACCACCGCGA CCTGGGTGCG CCACTACGTG GCCACGCGCG AGAAGAAAGT CAACGCTTC 960  
 ACCCAACTGCG TCAGCTGTCT GACTTTCGAA CTGAATATACC AGGTGCGGCT GTTGCGCGAC 1020  
 GAGAAATCTC TGAAGGCGCG CAAGGTGTGT TCTTACAGT CCGGCGTGTCT GAGAGGCAAG 1080  
 GACATGAGCA AGGTGTTCGC CTGATATGCG CCCAAGACCC TGATCTGAAA CTACTGGGTTC 1140  
 AACAACTACC TGCTCGGCA CACGCGCGCG GCCTTCGACA TCTCTTACTG GAAACAAGAC 1200  
 ACCACGCGCC TGCCCGCGCG GCTGACGCGC GAGTTCTGTG AACTGTTCAA GAGCAACCGC 1260  
 CTGAACCGCG CCGGCGCGCT GAGGTCTCC GGCACGCGCA TCGACTGAAA GCAAGTGAAT 1320  
 TGCGACTTCT ACTGTGTGCG CGGTCTGAAC GACCAATACA CCCCCTGGGA GTCTGTCTAC 1380  
 AAGTGGGCA GGTGTGTGCG TGCGAATGCG GAGTTCAATC TCTCCACAG CGGTCAATAC 1440  
 CAGAGCTCC TCACCCACC GGGCAACCC AAGGACGCT TCATGACCAA TCCGAAACTG 1500  
 CCGCGCGAG CCAAGGCTTG GCTGAAACAG GCCGCGACG AGCGGACTC GTGTGGTTG 1560  
 CACTGGCAG AATGGCTGCG GAAAGCTCC GGGAAACCC GCAAGGCGCC CGCGAGCTG 1620  
 GGCACAAAG CCTATCCGCG CGCGAAGCC GCGCGCGAA CCTAGTGTCA TGAACGA 1677

## (2) INFORMATION FOR SEQ ID NO: 16:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1680 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

AATGCTGAAA AGCAGGAATC GGGTAGCGTG CCGGTGCGCG CCGAGTTCAAT GAGTGACAG 60  
 AAGGCAATCG TGAGCTGCG CGGCAAGAGC CTGCTGACGA CGGTCCGCGAG CTGAGCTGTG 120  
 CACGCGCTGC GCGAGCGGCT GCAAGTGTG CGGACCTGTG TGCGCTTGTG AGGCCAAGTTG 180  
 GGCAGAGTGC TGCTGGGCGA CACCTGCGAC CAGCCGAACC CAAGAGAGCG CGGCTTCGAG 240  
 GATTCATCTT GGGCGCTTCA TCCCTTCTAC CGGCGACCC TGCAAGGCTTA CTGGGCGTGG 300  
 CAGAAACAC TGCTGCGCTG GATGAGAGAA AGCAACCTG ACTGACAGCA TCGGCGCGCG 360  
 GCGCGCTTC TGATGCTT GCTTCCGAC GCGGTGAGAC CCAAGCAAGC CTGATCAAT 420  
 CCACTGGCGT TAAAGAACT GTTCAATACC GCGGGAATCA GCTGTCTCAA TGCGCTCGCG 480

CACTGTCTG AAGACTGTGT GCAACAAGCG GCGATGCGCA GCGAGTGAAA CAAGACCGCC 540  
 TTGAGATCG GTGCAAACT CGCCACAGAG CAAGCGCGCG TGATGTTCG CAAGAGGTG 600  
 CTGAGACTGA TCGAGTACA GCGCTGCGCG GAGCGCAAGT AGCGCAAGC CCTGTCTATC 660  
 GTGCGCGCGC AGATCAACAA GTACTACATC TTGAACTGT CGCGGAAAA GAGTTTGTG 720  
 CAGTACGCGC TGAAGAACAA CTTGCAAGTC TTGATATCA GTTGGCGGCA CCGCGAGCGC 780  
 CAGCACGCGC AATGGGCGCT GAGCACTAT GTGAGGCGCC TCGAACAGAC CATGAGATC 840  
 AGCGCGAGAA TCAACGCGAG CCGGAGCTG AACCTGCGCG GCGCTGTGAG CCGCGGCGTC 900  
 ACCGTAGCGC CTTGTGTGCG CCACTGCGAG GTGCGCGCGC AACTGCGGAA GTTCAGTACG 960  
 GTCACTTACC TGATCAAGCT GCTGACAGC CAGATGAAA GCGCGGAGAT GCTTTGCGC 1020  
 GACGAGCAGA CCTTGAGAGAG CAGCAAGCGC GCTCTCTACC AGCATGCGGT GCTGACGCGG 1080  
 CGGACATCG CCAAGGTGTG CGCTGAGATG GCGCCAGAG ACCTGATCTG GAACTACTG 1140  
 GTCAACAAT ACTGTCTGCG CAGGCAAGCG CCGGCGTTG ACATCTCTTA CTGAAACAC 1200  
 GACACAGCG GGTGCGCGCG GCGTTCCAC GCGCAACTGC TCGAATCTGT CAAGACAGAC 1260  
 CCGTGAACCC GCGCGGCGCG GCTGGAAGTC AGCGGAGCG CGGTGAACTT GGGCAAGGTG 1320  
 GCGATGACA GCTTCACAT CGCGGAGATC ACCGACACA TCAGGCGCTG GAGCGGCTG 1380  
 TATGCTGCG CCGTCTGTCT GGGGAGCGAG CGCGCTTCA TCTGTTCGA CAGCGGCGAC 1440  
 ATCGAGAGCA TCTCAATCC TCCGGAAC CCCTAAGGCTT GCTACTTGA GAAAGCAAG 1500  
 CTGAGCAGCG ATCCAGCGCG CTGTACTAC GACGCGACG GCGAAGAGG CAAGTGTG 1560  
 CCGGTCTGCG TGAGCTGAGT GCAAGAGCGC TCGGCTGAGC TGCGCAACCC TGAATTCAAC 1620  
 CTGGCAGCG CCGCGATCC GCGCTGAAA GCGCGCGCG GCACTTACGT GCATATACG 1680

## (2) INFORMATION FOR SEQ ID NO: 17:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1791 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

ATGAGCCGAA AGAACAATTA CAGCTTCC AAGCAAGCG CGAATAACAC GCTAACTG 60

5  
AATCCGTTGA TCGCATCCG GGGCAAGGAC CTGCTACCT CCGCGCGCAT GGTCTGCTC 120  
CAGCGGTGCG CCACGCCGCT GCGACGCC AGGCGAGTGG CCGATTTCAG CTTGGAGCTG 180  
AAGAACGTCC TGCTCGGCCA GTCCGAGCTA CGCCCGAGCG ATGACGACCG AGCTTTTCC 240  
GATCCGGCTT GAGCCAGGAA TCCACTGTAC AAGCGCTACA TCCAGACCTA CTTGGCTGG 300  
10 CCGAAGAGC TGCACAGCTG GATCAGCCAC AGCGACCTGT CCGCGCAGGA CATCAGTCTT 360  
GGCCAGTTCC TCATCAACTT GCTGACCGAG GCGATGTCGC CGACCCACAG CTTGAGCACC 420  
CCGCGGCGCG TCACAGGCTT CTTGAGACC GCGCGCAGGA GCTTGTGGA CGGCTTCGCG 480  
15 CACCTGGCCA AGGACCTGCT GAACAACGCC GGGATGCCGA GCGAGGTGGA CATGAGACGC 540  
TTGAGGTGG CCAGAACCTT GGCACACACC GAGGGGCGCG TGGTGTTCGG CACGACGTG 500  
CTGGNACTGA TCCAGTACCG GCGCATCACC GAGTCGGTGC ACGAACGCC GCTGCTGGTG 560  
GTGCGGCGCG AGATCAACAA GTTCTACGTC TTCGACCTGT CCGCCGACAA GAGCTGGCG 720  
20 CGCTTCTGCC TGGCACAACG CGTGCAGACC TTCTATGCTCA GTTGGCGGAA CCGAACCAAG 780  
TCGACGCGCG AATGGGCTT GACCACTAT ATCGAGGCGC TCGAGGAGGC CATCGAGTGA 840  
GTCTGTGCGA TCACCGCGCAG CAGGACCTC AACCTCTCTG GCGCTGTCTC CGGGGGATC 900  
25 ACCACCGCGA CCCTGTTCCG CCACTACGTG GCCAGCGCG AGAAGAGGT CACGCGCTTC 960  
ACCCACTGG TCAGCTGCTT CAGCTTCGAA CTGATATACC AGTCTGCTT GTTCGCGCAC 1020  
GAGAGACTC TGGAGGCGCG CAGCGTCTT TCCTACCACT CCGCGCTGCT GGGGGCAG 1080  
35 GACATGGCCA AGGTGTTCCG CTGGATGCGC CCGACGACC TGATCTGGAA CTACTGGTTC 1140  
AACAACCTACC TGCTCGGCAA CCAGCGCGCG GCGTTCGACA TCTCTACTG GAACACGAC 1200  
ACCACGCGCC TGCCCGCGCG CTTGACGCGC GAGTTGCTCG AACTGTTCAA GAGCAACCG 1260  
40 CTGACACGCC CCGCGCGCTT GAGGTCTCC GGCACGCCCA TCGACCTGAA GCAGTGAAT 1320  
TGGGACTTCT ACTGTGTCGG CGGTCTGAAC GACCACATCA CCCCCTGGGA GTCTGTCTAC 1380  
AAGTCGGCCA GGTGCTGGGG TGGCAAGTGC GAGTTCAATC TTTCACACAG CGGTCAATC 1440  
CAGAGCATCC TCACCCACCC GGGCAACCCC AAGGCACGCT TCATGACCAA TCCGGAATG 1500  
50 CCGCGCGAGC CCAAGGCTTG GCTGGAAACAG GCGCGCAGC AGCGCACTC GTGGTGGTTG 1560  
CACTGCGAGC AATGGCTTGC CGAACGCTCC GGCAGACCC GCAAGGCGCC CGCCAGCCTG 1620  
GGCAACAGA CCTATCCGCG GCGCGAGGCC GCGCCCGGAA CTTAGTGTGA TCAAGCATCA 1680  
55 AAGCTTTGG GCAAGGTGT TACCGAGGA CAATTCABAG AGACCTGAC GAGGCCGGA 1740

## (2) INFORMATION FOR SEQ ID NO: 18:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 597 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS:  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

Met Ser Gln Lys Asn Asn Glu Leu Pro Lys Gln Ala Ala Glu Asn 15  
1 5 10  
Thr Leu Asn Leu Asn Pro Val Ile Gly Ile Arg Gly Lys Asp Leu Leu 30  
20 25  
Thr Ser Ala Arg Met Val Leu Leu Gln Ala Val Arg Gln Pro Leu His 45  
35 40  
Ser Ala Arg His Val Ala His Phe Ser Leu Glu Leu Lys Asn Val Leu 60  
50  
Leu Gly Gln Ser Glu Leu Arg Pro Gly Asp Asp Arg Arg Phe Ser 80  
65 70 75  
Asp Pro Ala Trp Ser Gln Asn Pro Leu Tyr Lys Arg Tyr Met Gln Thr 95  
85 90  
Tyr Leu Ala Trp Arg Lys Glu Leu His Ser Trp Ile Ser His Ser Asp 110  
100 105  
Leu Ser Pro Gln Asp Ile Ser Arg Gly Gln Phe Val Ile Asn Leu Leu 125  
115 120  
Thr Glu Ala Met Ser Pro Thr Asn Ser Leu Ser Asn Pro Ala Ala Val 140  
130 135  
Lys Arg Phe Phe Glu Thr Gly Lys Ser Leu Leu Asp Gly Leu Gly 160  
145 150 155  
His Leu Ala Lys Asp Leu Val Asn Asn Gly Gly Met Pro Ser Gln Val 175  
165 170  
Asp Met Asp Ala Phe Glu Val Gly Lys Asn Leu Ala Thr Thr Glu Gly 190  
180 185  
Ala Val Val Phe Arg Asn Asp Val Leu Glu Leu Ile Gln Tyr Arg Pro 205  
195 200

11e Thr Glu Ser Val His Glu Arg Pro Leu Leu Val Val Pro Pro Gln  
 210 215 220  
 11e Asn Lys Phe Tyr Val Phe Asp Leu Ser Pro Asp Lys Ser Leu Ala  
 225 230 235 240  
 Arg Phe Cys Leu Arg Asn Gly Val Gln Thr Phe Ile Val Ser Trp Arg  
 245 250 255  
 Asn Pro Thr Lys Ser Gln Arg Glu Trp Gly Leu Thr Thr Tyr Ile Glu  
 260 265 270  
 Ala Leu Lys Glu Ala Ile Glu Val Val Leu Ser Ile Thr Gly Ser Lys  
 275 280 285  
 Asp Leu Asn Leu Leu Gly Ala Cys Ser Gly Gly Ile Thr Thr Ala Thr  
 290 295 300  
 Leu Val Gly His Tyr Val Ala Ser Gly Glu Lys Lys Val Asn Ala Phe  
 305 310 315 320  
 Thr Gln Leu Val Ser Val Leu Asp Phe Glu Leu Asn Thr Gln Val Ala  
 325 330 335  
 Leu Phe Ala Asp Glu Lys Thr Leu Glu Ala Ala Lys Arg Arg Ser Tyr  
 340 345 350  
 Gln Ser Gly Val Leu Glu Gly Lys Asp Met Ala Lys Val Phe Ala Trp  
 355 360 365  
 Met Arg Pro Asn Asp Leu Ile Trp Asn Tyr Trp Val Asn Asn Tyr Leu  
 370 375 380  
 Leu Gly Asn Gln Pro Pro Ala Phe Asp Ile Leu Tyr Trp Asn Asn Asp  
 385 390 395 400  
 Thr Thr Arg Leu Pro Ala Ala Leu His Gly Glu Phe Val Glu Leu Phe  
 405 410 415  
 Lys Ser Asn Pro Leu Asn Arg Pro Gly Ala Leu Glu Val Ser Gly Thr  
 420 425 430  
 Pro Ile Asp Leu Lys Gln Val Thr Cys Asp Phe Tyr Cys Val Ala Gly  
 435 440 445  
 Leu Asn Asp His Ile Thr Pro Trp Glu Ser Cys Tyr Lys Ser Ala Arg  
 450 455 460  
 Leu Leu Gly Gly Lys Cys Glu Phe Ile Leu Ser Asn Ser Gly His Ile  
 465 470 475 480  
 Gln Ser Ile Leu Asn Pro Pro Gly Asn Pro Lys Ala Arg Phe Met Thr  
 485 490 495  
 Asn Pro Glu Leu Pro Ala Glu Pro Lys Ala Trp Leu Glu Gln Ala Gly  
 500 505 510

-77-

Lys His Ala Asp Ser Trp Trp Leu His Trp Gln Gln Trp Leu Ala Glu  
 515 520 525  
 Arg Ser Gly Lys Thr Arg Lys Ala Pro Ala Ser Leu Gly Asn Lys Thr  
 530 535 540  
 Tyr Pro Ala Gly Glu Ala Ala Pro Gly Thr Tyr Val His Glu Arg Ser  
 545 550 555 560  
 Lys Ala Leu Gly Lys Gly Val Thr Glu Glu Gln Phe Lys Glu Thr Trp  
 565 570 575  
 Thr Arg Pro Gly Ala Ala Gly Met Gly Glu Gly Thr Ser Leu Val Val  
 580 585 590  
 Ala Lys Ser Arg Met  
 595

(2) INFORMATION FOR SEQ ID NO: 19:  
 (1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1794 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 19:  
 ATGCGTGAAGA AGCAGGAAATC GAGTAGCCTG CCGGTGCCCG CCGAGTTCAAT GAGTGACAG 60  
 AGCGCCATCG TCGGCTTCGCG CGGCAAGAGC CTGCTGAGCA CGGTGCCGAG CCTGAGCTGC 120  
 CAGGCGCTGC GCCAGCGGCT GCACAGTGC CGGCACTTCG TCGCTTCG AGGCCAGTTG 180  
 GGCAGAGTGC TGCTGGAGCGA CACCTTCGAC CAGCCGAACC CACAGAGCG CCGCTTCGAG 240  
 GATCAATCT GGGGCTCTAA TCCCTTTAC CGGCGACCC TGCAGGCTTA CCTGGAGTGG 300  
 CAGAAACAC TGCTGCCCTG GATCGACGAA AGCAACTCG ACTGACAGCA TCGGCGCCCG 360  
 GCCGCTTCG TCGTGCCTT GCTCTCCGAC GCCGTGGGAC CCGAGACAG CCTGATCAAT 420  
 CCACTGAGCT TAAAGGAACT GTTCAATACC GGGGAGATCA GCTTGCTCAA TGGGTCGCC 480  
 CACTGCTCG AAGACTGCT GCACAAAGGC GGCATGCCCA GCCAGGTGAA CAAGACGCC 540  
 TTGAGATCG GTCCAACT CGCCACACAG CAAGCGGCG TGGTTCG CAAGAGTGG 600  
 CTGAGACTGA TCCAGTACAA GCCGTCGGGC GAGCGCCAGT AGGCCAAGCC CCTGCTGATC 660  
 GTGCGCGCG AGATCAACAA GTACTACATC TTGAACTGCT CCGCGAATAA GAGCTTCGTC 720

-78-



CAGTACGCC TGAAGAACAA CTTGCAGGTC TTCGTATCA GTTGGCGCAA CCCGACGCC 780  
 CAGCACCGC AATGGGGCT GAGCACTTAT GTGAGGGCC TCGACGAGC CATCGAGTC 840  
 AGCGCGAGA TCACCGGCG CCGACGCTG NACCTGGCCG GCGCTCGCG CGCGGGCTC 900  
 ACCGTAGCG CTTTGCTGG CCACCTGCAG GTGCGCGGC AACTCGCAA GGTCACTAGC 960  
 GTCACTTAC TGTGTAGCT GTTCGACAG CAGATGGAAA GCGCGGCGAT GCTCTGCGC 1020  
 GACGAGCAG CCCTGGGAG CACGAGCGC CGTCTTACC AGCATGGCT GCTGGAGGG 1080  
 CGCGACATGG CCAAGGTGT CGCTGGATG CGCCCAACG ACCTGATCTG GACTACTGG 1140  
 GTCAACAAT ACTGTCTGG CAGGCGGCG CGGCGGTTG ACATCTCTTA CTGGACACAC 1200  
 GACACACGC GGCTGCCCG GCGTTCCAC GCGGAACTGC TCGACCTGTT CAAGCACAC 1260  
 CGGCTGACCC GCGCGGCGC GCTGGAGGTC AGCGGGACCG CGGTGACCT GGGCAGGTC 1320  
 GCGATCGACA GTTTCCAGT CGCGGCGATC ACCGACACA TCACGCCCTG GAGCGGTCG 1380  
 TATCGCTCGG CCTTCTGCT GGGCGGCCAG CGCGCTTCA TCTGTCCAA CAGCGGGCAC 1440  
 ATCCAGAGA TCCTCAACCC TCCCGAAC CCGAGGCTT GCTACTTGA GACGACAG 1500  
 CTGAGCAGC ATCCAGCGC CTGTACTAC GAGCGCAAG CCGAGAGGG CAGCTGGTCG 1560  
 CGGCTCTGCG TGCGCTGCT GCAGGAGCG TCGGCGGAG TGGCGAAC TGACTTCAAC 1620  
 CTTGGCAGC CGCGCATCC GCCCTCGAA GCGGCGCGG GCNCTACGT GCATATACG 1680  
 TCAAAAGCTT TGGCGAAGG TGTACCGAG GAACAATCA AAGAGACCTG GACGAGCGC 1740  
 GAGCTGCTG GATGGGCGA AGGACTAGC CTTGTGTGG CCAAGTCCAG AATG 1794

## (2) INFORMATION FOR SEQ ID NO: 20:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 598 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS:

(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 20:

Met Arg Glu Lys Lys Glu Ser Gly Ser Val Pro Val Pro Ala Glu Phe  
 1 5 10 15  
 Met Ser Ala Ser Ala Ile Val Gly Leu Arg Gly Lys Asp Leu Leu  
 20 25 30

Thr Thr Val Arg Ser Leu Ala Val His Gly Leu Arg Gln Pro Leu His  
 35 40 45  
 Ser Ala Arg His Leu Val Ala Phe Gly Gly Gln Leu Gly Lys Val Leu  
 50 55 60  
 Leu Gly Asp Thr Leu His Gln Pro Asn Pro Gln Asp Ala Arg Phe Gln  
 65 70 75 80  
 Asp Pro Ser Trp Arg Leu Asn Pro Phe Tyr Arg Arg Thr Leu Gln Ala  
 85 90 95  
 Tyr Leu Ala Trp Gln Lys Gln Leu Ala Trp Ile Asp Glu Ser Asn  
 100 105 110  
 Leu Asp Cys Asp Asp Arg Ala Arg Ala Arg Phe Leu Val Ala Leu Leu  
 115 120 125  
 Ser Asp Ala Val Ala Pro Ser Asn Ser Leu Ile Asn Pro Leu Ala Leu  
 130 135 140  
 Lys Glu Leu Phe Asn Thr Gly Gly Ile Ser Leu Leu Asn Gly Val Arg  
 145 150 155 160  
 His Leu Leu Glu Asp Leu Val His Asn Gly Gly Met Pro Ser Gln Val  
 165 170 175  
 Asn Lys Thr Ala Phe Glu Ile Gly Arg Asn Leu Ala Thr Thr Gln Gly  
 180 185 190  
 Ala Val Val Phe Arg Asn Glu Val Leu Glu Leu Ile Gln Tyr Lys Pro  
 195 200 205  
 Leu Gly Glu Arg Gln Tyr Ala Lys Pro Leu Leu Ile Val Pro Pro Gln  
 210 215 220  
 Ile Asn Lys Tyr Tyr Ile Phe Asp Leu Ser Pro Glu Lys Ser Phe Val  
 225 230 235 240  
 Gln Tyr Ala Leu Lys Asn Asn Leu Gln Val Phe Val Ile Ser Trp Arg  
 245 250 255  
 Asn Pro Asp Ala Gln His Arg Glu Trp Gly Leu Ser Thr Tyr Val Glu  
 260 265 270  
 Ala Leu Asp Gln Ala Ile Glu Val Ser Arg Glu Ile Thr Gly Ser Arg  
 275 280 285  
 Ser Val Asn Leu Ala Gly Ala Cys Ala Gly Gly Leu Thr Val Ala Ala  
 290 295 300  
 Leu Leu Gly His Leu Gln Val Arg Arg Gln Leu Arg Lys Val Ser Ser  
 305 310 315 320  
 Val Thr Tyr Leu Val Ser Leu Leu Asp Ser Gln Met Glu Ser Pro Ala  
 325 330 335 340

	325	330	335
Met Leu Phe Ala Asp Glu Gln Thr Leu Glu Ser Ser Lys Arg Arg Ser	340	345	350
Tyr Gln His Gly Val Leu Asp Gly Arg Asp Met Ala Lys Val Phe Ala	355	360	365
Trp Met Arg Pro Asn Asp Leu Ile Trp Asn Tyr Trp Val Asn Asn Tyr	370	375	380
Leu Leu Gly Arg Gln Pro Pro Ala Phe Asp Ile Leu Tyr Trp Asn Asn	385	390	395
Asp Asn Thr Arg Leu Pro Ala Ala Phe His Gly Glu Leu Leu Asp Leu	405	410	415
Phe Lys His Asn Pro Leu Thr Arg Pro Gly Ala Leu Glu Val Ser Gly	420	425	430
Thr Ala Val Asp Leu Gly Lys Val Ala Ile Asp Ser Phe His Val Ala	435	440	445
Gly Ile Thr Asp His Ile Thr Pro Trp Asp Ala Val Tyr Arg Ser Ala	450	455	460
Leu Leu Leu Gly Gly Gln Arg Arg Phe Ile Leu Ser Asn Ser Gly His	465	470	475
Ile Gln Ser Ile Leu Asn Pro Pro Gly Asn Pro Lys Ala Cys Tyr Phe	485	490	495
Glu Asn Asp Lys Leu Ser Ser Asp Pro Arg Ala Trp Tyr Tyr Asp Ala	500	505	510
Lys Arg Glu Glu Gly Ser Trp Trp Pro Val Trp Leu Gly Trp Leu Gln	515	520	525
Glu Arg Ser Gly Glu Leu Gly Asn Pro Asp Phe Asn Leu Gly Ser Ala	530	535	540
Ala His Pro Pro Leu Glu Ala Ala Pro Gly Thr Tyr Val His Ile Arg	545	550	555
Ser Lys Ala Leu Gly Lys Gly Val Thr Glu Glu Gln Phe Lys Glu Thr	565	570	575
Trp Thr Arg Pro Gly Ala Ala Gly Met Gly Glu Gly Thr Ser Leu Val	580	585	590
Val Ala Lys Ser Arg Met	595		

(2) INFORMATION FOR SEQ ID NO: 21:

(1) SEQUENCE CHARACTERISTICS:

-81-

	(A) LENGTH: 2737 base pairs	
	(B) TYPE: nucleic acid	
	(C) STRANDEDNESS: single	
	(D) TOPOLOGY: linear	
	(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 21:	
GAATTCATGT CTCAGTTGA TTTAAAGAT AAGTTGTGA TCATTACCG TGCCGGTGT	60	
GGTTGGGGA AATACACTC CCTGCAATTT GCCAAGTTGG GGGCCAAAGT CGTGTTAAC	120	
GACTTGGGATG GTGCTTGGA CGGTCAAGGT GGAACTTCCA AGGCCGCCGA CGTGTGCGTT	180	
GACGAATTTG TCAGAAACGG TGGTGTGCG GTTGCCGAAAT ACAACAACGT CTTGACGGT	240	
GACAAATTTG TCGAACCGC CGTCAAGAC TTGGTACTG TCACAGTTAT CATCAACAT	300	
GGCGGTAATC TGAGAGATGC CTCGATGAG AAGATGACTG AAAAAGACTA CAATTGGTGC	360	
ATTGACGTGC ACTTGAAACGG TGCTTTGCG CTCACAAAGG CTGCTTGCC ATACTTCCA	420	
AAGCAAAAT ACAGTAATAT TGTCAACCA TCCTCCCGAG CTGTTTGTG CCGTAATTT	480	
GGTCAACCA ACTACCGCTC CGCAAGTCT GCTTTTGG GATTCGCTGA AACCTGGCC	540	
AAGAAAGTGG CCAATACCA CATCAAGCC AACGCCATGG CTCGTTGGC CAGATCAAGA	600	
ATGACTGAAT CTATCTTGC ACCCTCCATG TTGAAAAAT TGGGCTTGA AAAGTTGCG	660	
CAATTGCTCT TGTATTGTC GTGAGCTGA AACGAATGA CTGCGAATTT CTTGAAGTT	720	
GGTGTGGCT TTTACGCTCA GATCAGATGG GAAAGATCG GTGTGTCTT GTTCACCCA	780	
GATCAATCTT TCACCGCTGA GGTGTGCT AGAGATYCT CTGAATCTT TGATTATGAC	840	
GACTTAGGA AGCCAGATA CTTGAAGAC CAATACCAT TCAGTTGAA CGACTACGC	900	
ACTTGAACA ACAGACTAG AAAGTTGCCA GCTAACGATG CTTCTGGTGC TCACAATGTC	960	
TGCTTGAAG ACAAGTTGT TTGATACAC GGTGCCGGTG CTGTTTGGG TAAAGAAATAC	1020	
GCCAAATGAT TGCCCAAGTA CGGTGCAAG GTTGTGTGA ACGACTTCA GATGTACAC	1080	
AAGACGTTG ACGAATCAA AGCCGTGCT GGTGAAGCTT GGCAGATCA ACAAGATTT	1140	
GCCAAAGACT CCGAAGTAT CATCAAGAT GTCAATGACA AGTACGGTAC CATGATATC	1200	
TGAGTCAACA ACGCCGGTAT CTTGAGAGAC AGATCCTTGG CCAGATGTC CAAGCAAGA	1260	
TGGAGCTGAG TCCAAACAT CCACTGAAT GTTACTTTCA ACTGAGCAG ATTGGCAGG	1320	
CCATCTTGG TTGAAAACA ATTGGTAGA ATGATCAACA TTACTTCAC CAGTGTATC	1380	

-82-

5 TAGGTAACCT TTGGTCAAGC CAACACTACTCG TCTTCTAAGG CTGGTAATCTT GGGTTTGTCC 1440  
AAGACCATGG CCATTGAGGG TGTGTAAGAT AACATTAAAGG TCACATATGT TGCTCCACAC 1500  
GCTGMAACTG CCATGACCTT GACCATCTTC AGAGAACAAG ACAGAACTT GTACCACGCT 1560  
GACCAAGTTG CTCATTTGTT GGTCTACTTG GGTACTGACG ATGTCCCACT CACCGGTGAA 1620  
10 ACTTCCGAAA TCGGTGGTGG TTGGATCGGT AACACCAAGT GCGAAGAGC CAGGGTGTCT 1680  
GTCTCCCAAG ACAGACACAC CACTGTTGAA TTCACTAAGG AGCACTTGA CCAATCACT 1740  
GACTTCACCA CTGACACTGA AAATCCAAA TCTACCACCG AATCTCCAT GGTATCTTTG 1800  
15 TCTGCCGTTG GTGGTGATGA CGATGATGAT GACGAGAGC AAGAAGAGA CBAAGGTGAT 1860  
GAGAGAGAG ACAGAGAGA CAGAGAGAA GACGATCCAG TCTGAGATT CGAGACAGA 1920  
20 GATGTTATCT TGTACAACAT TGCCCTTGGT GCCACCACCA AGCAATTGAA GTACGCTTAC 1980  
GAAACGACT CTGACTTCCA AGTCATTCCA ACCTTTGGTC ACTTGATCAC CTTCAACTCT 2040  
GGTAAGTCAC AAACTCTCTT TGCCAAGTTG TTGGTAACT TCAACCCAAT GTTGTGTTG 2100  
25 CACGGTGAAC ACTACTTGA GGTGCACAGC TGGCCACAC CAAACGAGG TGAATCAAG 2160  
ACCACTTTCG AACCAATTGC CACTACTCCA AAGGGTACCA AGCTTGTAT TGTTCACGCT 2220  
TCCAAATCTG TTGACAACAA GTCTGGTGA TTGATTACT CCAACGAGC CACTTACTTC 2280  
ATCGAATCT GTCAACCGA CAAACAGTC TACGCTGACC GTCCAGCATT CGCCACCAAC 2340  
35 CAATTCTTGG CACCAAGAG AGCCCCAGC TACCAAGTTG AGCTTCCAGT CAGTGAAGAC 2400  
TTGGCTGCTT TGTACCGTTT GTCTGGTGAC AGAAACCCAT TGCACATTGA TCCAAACTTT 2460  
GCTAAGGTTG CCAAGTTCCC TAAGCCAAATC TTACACGGTA TGTGCACTTA TGGTTTGTAGT 2520  
40 GCTAAGGCTT TGATTGACAA GTTTGGTATG TTCACGAAA TCAAGGCCAG ATTCCCGGT 2580  
ATTGCTTCC CAGGTGAAC CTTGAGAGTC TTGGCATGGA AGGAAGCGA TGACACTATT 2640  
GTCTTCCAAA CTCATGTTGT TGAATAGGT ACTATTGCCA TTAACAACGC TGCTATTAAAG 2700  
45 TTAGTCGGTG ACAGACAAA GATCTAATGA AGGATCC 2737

## (2) INFORMATION FOR SEQ ID NO: 22:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 906 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS:  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 22:

5 Met Ser Pro Val Asp Phe Lys Asp Lys Val Val Ile Ile Thr Gly Ala 15  
1 Met Ser Pro Val Asp Phe Lys Asp Lys Val Val Ile Ile Thr Gly Ala 15  
Gly Gly Leu Gly Lys Tyr Tyr Ser Leu Glu Phe Ala Lys Leu Gly 30  
25 Gly Gly Leu Gly Lys Tyr Tyr Ser Leu Glu Phe Ala Lys Leu Gly 30  
Ala Lys Val Val Val Asn Asp Leu Gly Gly Ala Leu Asn Gly Gln Gly 45  
35 Ala Lys Val Val Val Asn Asp Leu Gly Gly Ala Leu Asn Gly Gln Gly 45  
15 Gly Asn Ser Lys Ala Ala Asp Val Val Val Asp Glu Ile Val Lys Asn 60  
50 Gly Asn Ser Lys Ala Ala Asp Val Val Val Asp Glu Ile Val Lys Asn 60  
Gly Gly Val Ala Val Ala Asp Tyr Asn Asn Val Leu Asp Gly Asp Lys 75  
65 Gly Gly Val Ala Val Ala Asp Tyr Asn Asn Val Leu Asp Gly Asp Lys 75  
20 Ile Val Glu Thr Ala Val Lys Asn Phe Gly Thr Val His Val Ile Ile 90  
85 Ile Val Glu Thr Ala Val Lys Asn Phe Gly Thr Val His Val Ile Ile 90  
Asn Asn Ala Gly Ile Leu Arg Asp Ala Ser Met Lys Lys Met Thr Glu 110  
100 Asn Asn Ala Gly Ile Leu Arg Asp Ala Ser Met Lys Lys Met Thr Glu 110  
Lys Asp Tyr Lys Leu Val Ile Asp Val His Leu Asn Gly Ala Phe Ala 125  
115 Lys Asp Tyr Lys Leu Val Ile Asp Val His Leu Asn Gly Ala Phe Ala 125  
30 Val Thr Lys Ala Ala Trp Pro Tyr Phe Gln Lys Gln Lys Tyr Gly Arg 140  
130 Val Thr Lys Ala Ala Trp Pro Tyr Phe Gln Lys Gln Lys Tyr Gly Arg 140  
Ile Val Asn Thr Ser Ser Pro Ala Gly Leu Tyr Gly Asn Phe Gly Gln 155  
145 Ile Val Asn Thr Ser Ser Pro Ala Gly Leu Tyr Gly Asn Phe Gly Gln 155  
35 Ala Asn Tyr Ala Ser Ala Lys Ser Ala Leu Leu Gly Phe Ala Glu Thr 175  
165 Ala Asn Tyr Ala Ser Ala Lys Ser Ala Leu Leu Gly Phe Ala Glu Thr 175  
Leu Ala Lys Glu Gly Ala Lys Tyr Asn Ile Lys Ala Asn Ala Ile Ala 190  
180 Leu Ala Lys Glu Gly Ala Lys Tyr Asn Ile Lys Ala Asn Ala Ile Ala 190  
Pro Leu Ala Arg Ser Arg Met Thr Glu Ser Ile Leu Pro Pro Met 205  
195 Pro Leu Ala Arg Ser Arg Met Thr Glu Ser Ile Leu Pro Pro Met 205  
40 Leu Glu Lys Leu Gly Pro Glu Lys Val Ala Pro Leu Val Leu Tyr Leu 220  
210 Leu Glu Lys Leu Gly Pro Glu Lys Val Ala Pro Leu Val Leu Tyr Leu 220  
Ser Ser Ala Glu Asn Glu Leu Thr Gly Gln Phe Phe Glu Val Ala Ala 240  
225 Ser Ser Ala Glu Asn Glu Leu Thr Gly Gln Phe Phe Glu Val Ala Ala 240  
50 Gly Phe Tyr Ala Gln Ile Arg Trp Glu Arg Ser Gly Gly Val Leu Phe 255  
245 Gly Phe Tyr Ala Gln Ile Arg Trp Glu Arg Ser Gly Gly Val Leu Phe 255  
Lys Pro Asp Gln Ser Phe Thr Ala Glu Val Val Ala Lys Arg Phe Ser 270  
260 Lys Pro Asp Gln Ser Phe Thr Ala Glu Val Val Ala Lys Arg Phe Ser 270

W09935278 PCT/US98/00083

Glu Ile Leu Asp Tyr Asp Asp Ser Arg Lys Pro Glu Tyr Leu Lys Asn  
275 280 285

Gln Tyr Pro Phe Met Leu Asn Asp Tyr Ala Thr Leu Thr Asn Glu Ala  
290 295 300

Arg Lys Leu Pro Ala Asn Asp Ala Ser Gly Ala Pro Thr Val Ser Leu  
305 310 315 320

Lys Asp Lys Val Val Leu Ile Thr Gly Ala Gly Ala Gly Leu Gly Lys  
325 330 335

Glu Tyr Ala Lys Trp Phe Ala Lys Tyr Gly Ala Lys Val Val Asn  
340 345 350

Asp Phe Lys Asp Ala Thr Lys Thr Val Asp Glu Ile Lys Ala Ala Gly  
355 360 365

Gly Glu Ala Trp Pro Asp Gln His Asp Val Ala Lys Asp Ser Glu Ala  
370 375 380

Ile Ile Lys Asn Val Ile Asp Lys Tyr Gly Thr Ile Asp Ile Leu Val  
385 390 395 400

Asn Asn Ala Gly Ile Leu Arg Asp Arg Ser Phe Ala Lys Met Ser Lys  
405 410 415

Gln Glu Trp Asp Ser Val Gln Gln Val His Leu Ile Gly Thr Phe Asn  
420 425 430

Leu Ser Arg Leu Ala Trp Pro Tyr Phe Val Glu Lys Gln Phe Gly Arg  
435 440 445

Ile Ile Asn Ile Thr Ser Thr Ser Gly Ile Tyr Gly Asn Phe Gly Gln  
450 455 460

Ala Asn Tyr Ser Ser Lys Ala Gly Ile Leu Gly Leu Ser Lys Thr  
465 470 475 480

Met Ala Ile Glu Gly Ala Lys Asn Asn Ile Lys Val Asn Ile Val Ala  
485 490 495

Pro His Ala Glu Thr Ala Met Thr Leu Thr Ile Phe Arg Glu Gln Asp  
500 505 510

Lys Asn Leu Tyr His Ala Asp Gln Val Ala Pro Leu Leu Val Tyr Leu  
515 520 525

Gly Thr Asp Asp Val Pro Val Thr Gly Glu Thr Ser Glu Ile Gly Gly  
530 535 540

Gly Trp Ile Gly Asn Thr Arg Trp Gln Arg Ala Lys Gly Ala Val Ser  
545 550 555 560

His Asp Glu His Thr Thr Val Glu Phe Ile Lys Glu His Leu Asn Glu  
565 570 575

-85-

W09935278 PCT/US98/00083

Ile Thr Asp Phe Thr Thr Asp Thr Glu Asn Pro Lys Ser Thr Thr Glu  
580 585 590

Ser Ser Met Ala Ile Leu Ser Ala Val Gly Gly Asp Asp Asp Asp  
595 600 605

Asp Glu Asp Glu Glu Glu Asp Glu Gly Asp Glu Glu Asp Glu Glu  
610 615 620

Asp Glu Glu Glu Asp Asp Pro Val Trp Arg Phe Asp Asp Asp Val  
625 630 635 640

Ile Leu Tyr Asn Ile Ala Leu Gly Ala Thr Thr Lys Gln Leu Lys Tyr  
645 650 655

Val Tyr Glu Asn Asp Ser Asp Phe Gln Val Ile Pro Thr Phe Gly His  
660 665 670

Leu Ile Thr Phe Asn Ser Gly Lys Ser Gln Asn Ser Phe Ala Lys Leu  
675 680 685

Leu Arg Asn Phe Asn Pro Met Leu Leu Leu His Gly Glu His Tyr Leu  
690 695 700

Lys Val His Ser Trp Pro Pro Thr Glu Gly Glu Ile Lys Thr Thr  
705 710 715 720

Phe Glu Pro Ile Ala Thr Thr Pro Lys Gly Thr Asn Val Val Ile Val  
725 730 735

His Gly Ser Lys Ser Val Asp Asn Lys Ser Gly Glu Leu Ile Tyr Ser  
740 745 750

Asn Glu Ala Thr Tyr Phe Ile Arg Asn Cys Gln Ala Asp Asn Lys Val  
755 760 765

Tyr Ala Asp Arg Pro Ala Phe Ala Thr Asn Gln Phe Leu Ala Pro Lys  
770 775 780

Arg Ala Pro Asp Tyr Gln Val Asp Val Pro Val Ser Glu Asp Leu Ala  
785 790 795 800

Ala Leu Tyr Arg Leu Ser Gly Asp Arg Asn Pro Leu His Ile Asp Pro  
805 810 815

Asn Phe Ala Lys Gly Ala Lys Phe Pro Lys Pro Ile Leu His Gly Met  
820 825 830

Cys Thr Tyr Gly Leu Ser Ala Lys Ala Leu Ile Asp Lys Phe Gly Met  
835 840 845

Phe Asn Glu Ile Lys Ala Arg Phe Thr Gly Ile Val Phe Pro Gly Glu  
850 855 860

Thr Leu Arg Val Leu Ala Trp Lys Glu Ser Asp Asp Thr Ile Val Phe  
865 870 875

-86-

865 870 875 880  
 Gln Thr His Val Val Asp Arg Gly Thr Ile Ala Ile Asn Ala Ala  
 885 890 895  
 5 Ile Lys Leu Val Gly Asp Lys Ala Lys Ile  
 900 905

(2) INFORMATION FOR SEQ ID NO: 23:  
 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 2737 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

10

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 23:

20

25 GGATCCCATGT CTCAGTTGA TTTTAAGAT AAAGTTGTGA TCATTACGG TGCCTGGTGT 60  
 25 GGTGTTGGTA AATACTACT CCTCGAATTT GCCAAGTTGG GCGCCAAAGT CGTGGTTAAC 120  
 GACTTGGGTG GTGCTTTGAA CGGTCAAGGT GGAATCTCCA AGGCGCGCGA CGTTGTGCTT 180  
 30 GACGAATTTG TCAGAACCG TGTGTTGCC GTTGCAGATT ACACAACGT CTTGGACGGT 240  
 GACAGATTGG TCGAACCOC GGTCAAGAC TTTGGTACTG TCCAGTTAT CATCAACAT 300  
 GCGGTATCT TGAGAGATGC CTCATGAG AGATGACTG AAAAGACTA CAATTTGGTC 360  
 35 ATTGACGTGC ACTTGNACGG TGCTTTTGGC GTCACCAAGG CTGCTTGCC ATACTTCCAA 420  
 AAGCAAAAT ACGGTAGAT TGTCAACACA TCCTCCCGAG CTGGTTTGTGTA CGGTACTTT 480  
 GGTCAAGCCA ACTACGCTC CGCAAGTCT GCTTTGTTGG GATTCGCTGA AACCTTGGCC 540  
 40 AAGGAAGTG CCAAAATACAA CATCAAGGCC AAGGCCATTT CTCCGTTGGC CAGATCAAGA 600  
 ATGACTGAAT CTATCTTGGC ACCTCCATG TTGMAAAAT TGGGCCCTGA AAAGTTGCC 660  
 45 CAAATGGTCT TGTATTGTC GTACGTGAA AACGAATTGA CTGGTCAAT CTTTGAAGTT 720  
 GCTGCTGGCT TTAGGCTCA GATCAGATGG GAAAGATCCG GTGGTGTCTT GTTCAAGCCA 780  
 GATCAATCTT TCACCGCTGA GGTGTGTGCT AAGAGATCT CTGAATCTCT TGAATATGAC 840  
 50 GACTCTAGGA AGCCAGATA CTTGAAGAAC CAATACCCAT TCAATTGAA CCACTACGCC 900  
 ACTTTGACCA ACGAAGCTAG AAAGTTGCCA GCTAAGATG CTTCTGGTGC TCCAACTGTC 960  
 55 TCCTTGAAG ACAAGTTGT TTTGATCACC GGTGCGGTG CTGGTTTGGG TAAAGATAC 1020

GCCAAATGTT TCCCAAGTA CGTGCCNAG GTTGTGTTA AGGACTTCAA GGATGCTACC 1080  
 AAGACGGTTG ACGAATCAA AGCCGCTGGT GGTGAAGCTT GGCCAGATCA ACAGATGTT 1140  
 5 GCCAAGGACT CCGAAGCTAT CATCAAGAT GTCAATTGACA AGTACGGTAC CAATTGATATC 1200  
 TTGGTCAACA ACGCCGGTAT CTTGAGAGAC AGATCCCTTTG CCAAGATGTC CAGCAGAA 1260  
 TGGGACTCTG TCCNACAGT CCATTGATTTT GGTACTTTCA ACTTGAGCAG ATTGGCATGG 1320  
 10 CCATACTTTG TTGMAAACA ATTGTGTAGA ATCATCAACA TTACCTCCAC CAGTGGTATC 1380  
 TACGGTAAT TTTGTCNAGC CAATCTATCG TCTTTAAGG CTGGTATCTT GGGTTTGTCC 1440  
 15 AAGACCATGG CCATTGAAGG TGTGAAGAT AACATTAAAG TCAACATTGT TGCTCCNAC 1500  
 GCTGMAACTG CCATGACCTT GACCATCTTC AGAGAACAAG ACAAGAACTT GTACCACGCT 1560  
 GACCAAGTTG CTCATTTGTT GGTCTACTTG GGTACTGACG ATGTCCAGT CACCGGTGAA 1620  
 20 ACTTCCGAAA TCGGTGTGTG TTGGATCGGT AACACCAGAT GGCAAGAGC CAAGGTGCT 1680  
 GTCTCCGAG ACGACACAC CACTGTGGA TTCACTAAGG AGCACTTGA CCAATCACT 1740  
 25 GACTTCACA CTGACACTGA AATCCAAA TCTACCACCG AATCTCCAT GGTATCTTG 1800  
 TCTCCGTTG GTGGTGATGA CGATGATGAT GACGAAGACG AAGAAGAAGA CCAAGGTGAT 1860  
 GAAGAAGAG ACGAAGAGA CGAAGAGA GACGATCCAG TCTGGAGATT CGACGACAGA 1920  
 30 GATGTTATCT TGTACAACT TGCCTTGGT GCCACCACCA AGCAATTGAA GTACGTCTAC 1980  
 GAAAGGACT CTGACTTCCA AGTCATTCCA ACCTTTGGTC ACTTGATCAC CTTCAACTCT 2040  
 35 GGTAACTAC AAAACTCTCT TGCCTAAGTTG TTGCTAACT TCAACCCAAAT GTTGTGTTG 2100  
 CACGGTGAC ACTACTTGA GGTGCACAGC TGGCCACCAC CAACCGAAG TGAATCAG 2160  
 ACCACTTTC AACCAATTGC CACTACTCCA AAGGTACCA ACGTTGTTAT TGTTCAGGT 2220  
 40 TCCAAATCTG TTGACAACA GTCTGTGAA TTGATTACT CCAACGAAGC CACTTACTTC 2280  
 ATCAGAACT GTCAAGCGA CAACAAGTC TACGCTGACC GTCCAGCAAT GGCACCCAC 2340  
 45 CAATCTCTG CACCAAGAG AGCCCGAC TACCAGTTG ACGTTCCAGT CAGTGAAGAC 2400  
 TTGGCTGCTT TGTACCGTTT GTCTGTGAC AGAAACCCAT TGCACATTGA TCCAACTTT 2460  
 GCTAAAGTG CCAAGTTCCC TAAGCCAATC TTACAGGTA TGTGCACTTA TGGTTGAGT 2520  
 50 GCTAAGGCT TGAATGACA GTTTGGTATG TTCAACGAAA TCAAGGCCAG ATTCAACGGT 2580  
 ATTGTCTCC CAGGTGAAC CTTGAGATC TTGGCATGGA AGGAAGCGA TGACACTATT 2640  
 55 GTCTTCCAAA CTCATGTGT TGAATAGGT ACTATTGCCA TTAACACGC TGCTATTAG 2700

TTAGTGGTG ACAAAATCCAA GTTGTATATGA AGCATCC

2737

## (2) INFORMATION FOR SEQ ID NO: 24:

## 5 (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 906 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS:  
(D) TOPOLOGY: linear

10

## 15 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 24:

Met Ser Pro Val Asp Phe Lys Asp Lys Val Val Ile Ile Thr Gly Ala  
1 5 10 15  
Gly Gly Gly Leu Gly Lys Tyr Ser Leu Glu Phe Ala Lys Leu Gly  
20 25 30  
Ala Lys Val Val Asn Asp Leu Gly Gly Ala Leu Asn Gly Gln Gly  
35 40 45  
Gly Asn Ser Lys Ala Ala Asp Val Val Asp Glu Ile Val Lys Asn  
50 55 60  
Gly Gly Val Ala Val Ala Asp Tyr Asn Asn Val Leu Asp Gly Asp Lys  
65 70 75 80  
Ile Val Glu Thr Ala Val Lys Asn Phe Gly Thr Val His Val Ile Ile  
85 90 95  
Asn Asn Ala Gly Ile Leu Arg Asp Ala Ser Met Lys Lys Met Thr Glu  
100 105 110  
Lys Asp Tyr Lys Leu Val Ile Asp Val His Leu Asn Gly Ala Phe Ala  
115 120 125  
Val Thr Lys Ala Ala Trp Pro Tyr Phe Gln Lys Gln Lys Tyr Gly Arg  
130 135 140  
Ile Val Asn Thr Ser Ser Pro Ala Gly Leu Tyr Gly Asn Phe Gly Gln  
145 150 155 160  
Ala Asn Tyr Ala Ser Ala Lys Ser Ala Leu Leu Gly Phe Ala Glu Thr  
165 170 175  
Leu Ala Lys Glu Gly Ala Lys Tyr Asn Ile Lys Ala Asn Ala Ile Ala  
180 185 190  
Pro Leu Ala Arg Ser Arg Met Thr Glu Ser Ile Leu Pro Pro Met  
195 200 205  
Leu Glu Lys Leu Gly Pro Glu Lys Val Ala Pro Leu Val Leu Tyr Leu

-89-

Ser Ser Ala Glu Asn Glu Leu Thr Gly Gln Phe Phe Glu Val Ala Ala  
210 215 220 225 230 235 240  
Gly Phe Tyr Ala Gln Ile Arg Trp Glu Arg Ser Gly Gly Val Leu Phe  
245 250 255  
Lys Pro Asp Gln Ser Phe Thr Ala Glu Val Val Ala Lys Arg Phe Ser  
260 265 270  
Glu Ile Leu Asp Tyr Asp Asp Ser Arg Lys Pro Glu Tyr Leu Lys Asn  
275 280 285  
Gln Tyr Pro Phe Met Leu Asn Asp Tyr Ala Thr Leu Thr Asn Glu Ala  
290 295 300  
Arg Lys Leu Pro Ala Asn Asp Ala Ser Gly Ala Pro Thr Val Ser Leu  
305 310 315 320  
Lys Asp Lys Val Val Leu Ile Thr Gly Ala Gly Ala Gly Leu Gly Lys  
325 330 335  
Glu Tyr Ala Lys Trp Phe Ala Lys Tyr Gly Ala Lys Val Val Asn  
340 345 350  
Asp Phe Lys Asp Ala Thr Lys Thr Val Asp Glu Ile Lys Ala Ala Gly  
355 360 365  
Gly Glu Ala Trp Pro Asp Gln His Asp Val Ala Lys Asp Ser Glu Ala  
370 375 380  
Ile Ile Lys Asn Val Ile Asp Lys Tyr Gly Thr Ile Asp Ile Leu Val  
385 390 395 400  
Asn Asn Ala Gly Ile Leu Arg Asp Arg Ser Phe Ala Lys Met Ser Lys  
405 410 415  
Gln Glu Trp Asp Ser Val Gln Gln Val His Leu Ile Gly Thr Phe Asn  
420 425 430  
Leu Ser Arg Leu Ala Trp Pro Tyr Phe Val Glu Lys Gln Phe Gly Arg  
435 440 445  
Ile Ile Asn Ile Thr Ser Thr Ser Gly Ile Tyr Gly Asn Phe Gly Gln  
450 455 460  
Ala Asn Tyr Ser Ser Ser Lys Ala Gly Ile Leu Gly Leu Ser Lys Thr  
465 470 475 480  
Met Ala Ile Glu Gly Ala Lys Asn Asn Ile Lys Val Asn Ile Val Ala  
485 490 495  
Pro His Ala Glu Thr Ala Met Thr Leu Thr Ile Phe Arg Glu Gln Asp  
500 505 510

-90-

Lys Asn Leu Tyr His Ala Asp Gln Val Ala Pro Leu Leu Val Tyr Leu  
 515 520 525  
 Gly Thr Asp Asp Val Pro Val Thr Gly Glu Thr Ser Glu Ile Gly Gly  
 530 535 540  
 Gly Tyr Ile Gly Asn Thr Arg Trp Gln Arg Ala Lys Gly Ala Val Ser  
 545 550 555 560  
 His Asp Glu His Thr Thr Val Glu Phe Ile Lys Glu His Leu Asn Glu  
 565 570 575  
 Ile Thr Asp Phe Thr Thr Asp Thr Glu Asn Pro Lys Ser Thr Thr Glu  
 580 585 590  
 Ser Ser Met Ala Ile Leu Ser Ala Val Gly Gly Asp Asp Asp Asp  
 595 600 605  
 Asp Glu Asp Glu Glu Glu Asp Glu Gly Asp Glu Glu Asp Glu Glu  
 610 615 620  
 Asp Glu Glu Glu Asp Asp Pro Val Trp Arg Phe Asp Asp Arg Asp Val  
 625 630 635 640  
 Ile Leu Tyr Asn Ile Ala Leu Gly Ala Thr Thr Lys Lys Glu Lys Tyr  
 645 650 655  
 Val Tyr Glu Asn Asp Ser Asp Phe Gln Val Ile Pro Thr Phe Gly His  
 660 665 670  
 Leu Ile Thr Phe Asn Ser Gly Lys Ser Gln Asn Ser Phe Ala Lys Leu  
 675 680 685  
 Leu Arg Asn Phe Asn Pro Met Leu Leu Leu His Gly Glu His Tyr Leu  
 690 695 700  
 Lys Val His Ser Trp Pro Pro Thr Glu Gly Glu Ile Lys Thr Thr  
 705 710 715 720  
 Phe Glu Pro Ile Ala Thr Thr Pro Lys Gly Thr Asn Val Val Ile Val  
 725 730 735  
 His Gly Ser Lys Ser Val Asp Asn Lys Ser Gly Glu Leu Ile Tyr Ser  
 740 745 750  
 Asn Glu Ala Thr Tyr Phe Ile Arg Asn Cys Gln Ala Asp Asn Lys Val  
 755 760 765  
 Tyr Ala Asp Arg Pro Ala Phe Ala Thr Asn Gln Phe Leu Ala Pro Lys  
 770 775 780  
 Arg Ala Pro Asp Tyr Gln Val Asp Val Pro Val Ser Glu Asp Leu Ala  
 785 790 795 800  
 Ala Leu Tyr Arg Leu Ser Gly Asp Arg Asn Pro Leu His Ile Asp Pro  
 805 810 815

-91-

Asn Phe Ala Lys Gly Ala Lys Phe Pro Lys Pro Ile Leu His Gly Met  
 820 825 830  
 Cys Thr Tyr Gly Leu Ser Ala Lys Ala Leu Ile Asp Lys Phe Gly Met  
 835 840 845  
 Phe Asn Glu Ile Lys Ala Arg Phe Thr Gly Ile Val Phe Pro Gly Glu  
 850 855 860  
 Thr Leu Arg Val Leu Ala Trp Lys Glu Ser Asp Asp Thr Ile Val Phe  
 865 870 875 880  
 Gln Thr His Val Val Asp Arg Gly Thr Ile Ala Ile Asn Asn Ala Ala  
 885 890 895  
 Ile Lys Leu Val Gly Asp Lys Ser Lys Leu  
 900

(2) INFORMATION FOR SEQ ID NO: 25:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2737 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 25:

GGATCATGTCCTCCAGTTGTA TTTTAAAGAT AAGTTTGTA TCATTACCGG TGCCGGTGT 60  
 GGTTTGGGTA AATACTACTC CCTCGAATTT GCCAAGTTGG GCGCCAAAGT CGTCGTTAAC 120  
 GACTTGGGTG GTGCTTGAA CGGTCAAGGT GGAACACTCA AGGCCGCCGA CGTTGTCGTT 180  
 GACGAAATTT TCAGAACGG TGGTGTGGC GTTGCCGATT ACACACACGT CTGGACGCT 240  
 GACAGATTT TCGAAACCGC CGTCAAGAAC TTGTGTACTG TCCACGTTAT CATCAACAAAT 300  
 CCGGTAATTT TGAGAGATGC CTCCATGAAG AAGATGACTG AAAAAGACTA CAAATTGGTC 360  
 ATTGACGTGC ACTTGACGG TGCCTTTGGC GTCACCAAGG CTGCTTGGCC ATACTTCCAA 420  
 AAGCAAAAT ACGGTAGAAT TGTCACACA TCCTCCCCAG CTGGTTTGTA CGGTAACTTT 480  
 GGTCAAGCCA ACTACGCCTC CGCCAAGTCT GCTTTGTGG GATTCGCTGA AACCTTGGCC 540  
 AAGGAAGGTG CCAATACAA CATCAAGGCC AACGCCATTT CTCGGTTGGC CAGATCAGA 600  
 ATGACTGAAT CTATCTTCCC ACCTCCAATG TTGGAAAAAT TGGGCCCTGA AAGGTTGGC 660  
 CCATTGCTCT TGTATTGTC GTCAGCTGAA AACGAATTGA CTGGTCAATT CTTTGAAGTT 720

-92-

GCTGCTGACT TTAAAGCTCA GATCGATGAG GAAAGATCCG GTGTGTCTCT GTTCAAGCCA 780  
 GATCATTCCT TCACCGCTGA GGTGTGTCTT AAGGATTCCT CTGAATTCCT TGATTAATAC 840  
 5 GACTCTAGA ACCCGAATAT CTGAGAGAC CAATPACCAT TCATGTGTGA CGACTACGCC 900  
 ACTTTAGCA ACGAAGCTAG AAGTTGCCA GCTAAGCATG CTTCGTGTGC TCCAACTGTG 960  
 10 TCCTTGAGG ACAAGGTGT TTGATACG GTGTCCGGTG CTGTTTGGG TAAAGAAATC 1020  
 GCCAAGGTGT TCGCAAGTA CGGTGCCAG GTTGTGTGA ACGACTTCA GGAATGTACC 1080  
 AAGACCGTT ACGAATTCGA AGCCGCTGTG GTTGAACTT GCCCAATCA ACAAGATGT 1140  
 15 GCCAAGACT CCGAAGCTAT CATCAAGAT GTCAATGACA AGTACGGTAC CATTGATATC 1200  
 TTGCTCAACA ACGCCGGTAT CTGAGAGAC AGATCCCTTG CCAAGATGTC CAAGCAAGAA 1260  
 20 TGGGACTCTG TCCAAAGAT CCACTGAT GTTACTTTCA ACTGAGCAG ATTGGCATGG 1320  
 CCAATCTTGG TTGAAAAACA ATTGTGTGA ATCATACACA TTACCTTCAC CAGTGTATC 1380  
 TAGGTAAT TTGCTCAAGC CAATCTCTG TCTTCTAAG CTGGTATCT GGGTTGTGCC 1440  
 25 AAGACCAAG CCAITGAGG TGTAAAGAT AACATTAAG TCACATTTGT TGCTCCACAC 1500  
 GCTGAAGCT CCAATGACCT GACCATCTTC AGAGAGACAG ACGAAGACT GTACACGGCT 1560  
 30 GACCAAGTTG CTCATCTGTT GGTCTACTTG GGTACTGAGG ATGTCCAGG CACCGGTGAA 1620  
 ACTTCGAAA TCGGTGTGTG TTGATGTGT AACACAGAT GGCAGAGAG CAGGGTGTCT 1680  
 GTTCCCAAG ACGAAGACAC CACTGTGAA TTCAATCAGG AGCACTTGA CCAATCTACT 1740  
 35 GACTTCACA CTGACACTGA AATTCAGAAA TCTACACCG AATTCCTCAT GGTATCTTGG 1800  
 TCTGCCGTTG GTGTGTATGA CATTGATGT GACGAGAGAG AAGAGAGAG CAGAGGTGAT 1860  
 40 GAGAGAGAG ACGAGAGAGA CGAAGAGAA GACATCCAG TCTGAGATT CGACGACAGA 1920  
 GATGTATCT TGTACACAT TGCCTTGTG GCCACACCA AGCAATTTGA GTACGTCTAC 1980  
 45 GAAAGGACT CTGACTTCA AGTCATCTCA ACCTTGTGTC ACTTGATCAC CTTCACCTCT 2040  
 GGTAACTCAC AAAATCTCT TGCAGATTG TTGCGTAACT TCACCCCAT GTTGTGTGTTG 2100  
 CACGGTAAC ACTACTTGA GGTGACAGC TGGCCACAC CAACGAGG TGAATACAG 2160  
 50 ACGACTTGG AACCAATGTC CACTACTCA AAGGTACCA AGCTTGTAT TGTTCAGGT 2220  
 TCCAAATCTG TTACAGACA GTCTGTGAA TTGATTTACT CCAAGAGC CACTTACTTC 2280  
 ATCAAGAACT GTCAAGCGA CAAGAGGTC TACGCTGACC GTCCAGCAAT CGCCACCAAC 2340  
 55 CAATCTTGG CACCAAGAG AGCCCAAGC TACCAAGTTG AGCTTCAGT CAGTGAAGAC 2400

-93-

TTGGCTGCTT TGTACCTTT GTCTGTGAC AAGAAACCAT TGCATTTGA TCCAACTTT 2460  
 GCTAAGATG CCAAGTCCC TAAGCAATC TTACAGGTA TGTCACTTA TGTGTGAGT 2520  
 5 GCTAAGCTT TGAATGACA GTTGTGTATG TTCAAGAAA TCAAGCCAG ATTACCGGT 2580  
 ATGTCTTCC CAGGTAAAC CTGAGATC TTGGCATGGA AGAAGACCA TGAACATAT 2640  
 10 GTCTCCAAA CTGATGTGT TATAGAGT ACTATTCGA TTACAAAGC TGTATTTAAG 2700  
 TTAATCGGTG ACAATGAAA GATCAATGA AGATCC 2737

(2) INFORMATION FOR SEQ ID NO: 26:  
 (1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 903 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS:  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 26:  
 Met Ser Pro Val Asp Phe Lys Asp Lys Val Val Ile Ile Thr Gly Ala  
 1 5 10 15  
 Gly Gly Gly Leu Gly Lys Tyr Tyr Ser Leu Glu Phe Ala Lys Leu Gly  
 20 25 30  
 Ala Lys Val Val Val Asn Asp Leu Gly Gly Ala Leu Asn Gly Glu Gly  
 35 40 45  
 Gly Asn Ser Lys Ala Ala Asp Val Val Val Asp Glu Ile Val Lys Asn  
 50 55 60  
 Gly Gly Val Ala Val Ala Asp Tyr Asn Asn Val Leu Asp Gly Asp Lys  
 65 70 75 80  
 Ile Val Glu Thr Ala Val Lys Asn Phe Gly Thr Val His Val Ile Ile  
 85 90 95  
 Asn Asn Ala Gly Ile Leu Arg Asp Ala Ser Met Lys Lys Met Thr Glu  
 100 105 110  
 Lys Asp Tyr Lys Leu Val Ile Asp Val His Leu Asn Gly Ala Phe Ala  
 115 120 125  
 Val Thr Lys Ala Ala Tyr Pro Tyr Phe Glu Lys Glu Lys Tyr Gly Arg  
 130 135 140  
 Ile Val Asn Thr Ser Ser Pro Ala Gly Leu Tyr Tyr Asn Phe Gly Glu  
 145 150 155 160

-94-



Ala Asn Tyr Ala Ser Ala Lys Ser Ala Leu Leu Gly Phe Ala Glu Thr  
165 170 175  
5 Leu Ala Lys Glu Gly Ala Lys Tyr Asn Ile Lys Ala Asn Ala Ile Ala  
180 185 190  
Pro Leu Ala Arg Ser Arg Met Thr Glu Ser Ile Leu Pro Pro Met  
195 200 205  
10 Leu Glu Lys Leu Gly Pro Glu Lys Val Ala Pro Leu Val Leu Tyr Leu  
210 215 220  
Ser Ser Ala Glu Asn Glu Leu Thr Gly Gln Phe Glu Val Ala Ala  
225 230 235 240  
15 Gly Phe Tyr Ala Gln Ile Arg Trp Glu Arg Ser Gly Val Leu Phe  
245 250 255  
20 Lys Pro Asp Gln Ser Phe Thr Ala Glu Val Val Ala Lys Arg Phe Ser  
260 265 270  
Glu Ile Leu Asp Tyr Asp Ser Arg Lys Pro Glu Tyr Leu Lys Asn  
275 280 285  
25 Gln Tyr Pro Phe Met Leu Asn Asp Tyr Ala Thr Leu Thr Asn Glu Ala  
290 295 300  
Arg Lys Leu Pro Ala Asn Asp Ala Ser Gly Ala Pro Thr Val Ser Leu  
305 310 315 320  
Lys Asp Lys Val Val Leu Ile Thr Gly Ala Gly Ala Gly Leu Gly Lys  
325 330 335  
35 Glu Tyr Ala Lys Trp Phe Ala Lys Tyr Gly Ala Lys Val Val Val Asn  
340 345 350  
Asp Phe Lys Asp Ala Thr Lys Thr Val Asp Glu Ile Lys Ala Ala Gly  
355 360 365  
40 Gly Glu Ala Trp Pro Asp Gln His Asp Val Ala Lys Asp Ser Glu Ala  
370 375 380  
Ile Ile Lys Asn Val Ile Asp Lys Tyr Gly Thr Ile Asp Ile Leu Val  
385 390 395 400  
Asn Asn Ala Gly Ile Leu Arg Asp Arg Ser Phe Ala Lys Met Ser Lys  
405 410 415  
50 Gln Glu Trp Asp Ser Val Gln Gln Val His Leu Ile Gly Thr Phe Asn  
420 425 430  
Leu Ser Arg Leu Ala Trp Pro Tyr Phe Val Glu Lys Gln Phe Gly Arg  
435 440 445  
55 Ile Ile Asn Ile Thr Ser Thr Ser Gly Ile Tyr Gly Asn Phe Gly Gln  
450 455 460

450 455 460  
Ala Asn Tyr Ser Ser Lys Ala Gly Ile Leu Gly Leu Ser Lys Thr  
465 470 475 480  
Met Ala Ile Glu Gly Ala Lys Asn Asn Ile Lys Val Asn Ile Val Ala  
485 490 495  
Pro His Ala Glu Thr Ala Met Thr Leu Thr Ile Phe Arg Glu Gln Asp  
500 505 510  
Lys Asn Leu Tyr His Ala Asp Gln Val Ala Pro Leu Leu Val Tyr Leu  
515 520 525  
15 Gly Thr Asp Asp Val Pro Val Thr Gly Glu Thr Ser Glu Ile Gly Gly  
530 535 540  
Gly Trp Ile Gly Asn Thr Arg Trp Gln Arg Ala Lys Gly Ala Val Ser  
545 550 555 560  
20 His Asp Glu His Thr Thr Val Glu Phe Ile Lys Glu His Leu Asn Glu  
565 570 575  
Ile Thr Asp Phe Thr Thr Asp Thr Glu Asn Pro Lys Ser Thr Thr Glu  
580 585 590  
25 Ser Ser Met Ala Ile Leu Ser Ala Val Gly Asp Asp Asp Asp Asp  
595 600 605  
Asp Glu Asp Glu Glu Glu Asp Glu Gly Asp Glu Glu Asp Glu Glu  
610 615 620  
Asp Glu Glu Glu Asp Asp Pro Val Trp Arg Phe Asp Asp Arg Asp Val  
625 630 635 640  
35 Ile Leu Tyr Asn Ile Ala Leu Gly Ala Thr Thr Lys Gln Leu Lys Tyr  
645 650 655  
Val Tyr Glu Asn Asp Ser Asp Phe Gln Val Ile Pro Thr Phe Gly His  
660 665 670  
40 Leu Ile Thr Phe Asn Ser Gly Lys Ser Gln Asn Ser Phe Ala Lys Leu  
675 680 685  
Leu Arg Asn Phe Asn Pro Met Leu Leu His Gly Glu His Tyr Leu  
690 695 700  
45 Lys Val His Ser Trp Pro Pro Thr Glu Gly Glu Ile Lys Thr Thr  
705 710 715 720  
Phe Glu Pro Ile Ala Thr Thr Pro Lys Gly Thr Asn Val Val Ile Val  
725 730 735  
50 His Gly Ser Lys Ser Val Asp Asn Lys Ser Gly Glu Leu Ile Tyr Ser  
740 745 750

Asn Glu Ala Thr Tyr Phe Ile Arg Asn Cys Gln Ala Asp Asn Lys Val  
 755 760 765  
 Tyr Ala Asp Arg Pro Ala Phe Ala Thr Asn Gln Phe Leu Ala Pro Lys  
 770 775 780  
 Arg Ala Pro Asp Tyr Gln Val Asp Val Pro Val Ser Glu Asp Leu Ala  
 785 790 795 800  
 Ala Leu Tyr Arg Leu Ser Gly Asp Arg Asn Pro Leu His Ile Asp Pro  
 805 810 815  
 Asn Phe Ala Lys Gly Ala Lys Phe Pro Lys Pro Ile Leu His Gly Met  
 820 825 830  
 Cys Thr Tyr Gly Leu Ser Ala Lys Ala Leu Ile Asp Lys Phe Gly Met  
 835 840 845  
 Phe Asn Glu Ile Lys Ala Arg Phe Thr Gly Ile Val Phe Pro Gly Glu  
 850 855 860  
 Thr Leu Arg Val Leu Ala Trp Lys Glu Ser Asp Asp Thr Ile Val Phe  
 865 870 875 880  
 Gln Thr His Val Val Asp Arg Gly Thr Ile Ala Ile Asn Asn Ala Ala  
 885 890 895  
 Ile Lys Leu Val Gly Asp Lys  
 900

**WHAT IS CLAIMED IS:**

1. A non-naturally occurring fusion protein comprising:  
a peroxisome targeting protein subunit; and  
a polyhydroxyalkanoate synthase protein subunit.
2. The fusion protein of claim 1, wherein the peroxisome targeting subunit is PTS2.
3. The fusion protein of claim 1, wherein the peroxisome targeting subunit comprises a tripeptide, wherein:  
the first amino acid in the N-terminus to C-terminus direction is S, A, or P;  
the second amino acid in the N-terminus to C-terminus direction is K, R, S, or H;  
and  
the third amino acid in the N-terminus to C-terminus direction is L, M, I, or F.
4. The fusion protein of claim 3, wherein the peroxisome targeting subunit comprises ARM, SRM, SKL, ARL, SRL, PSI, or PRM.
5. The fusion protein of claim 1, wherein the peroxisome targeting subunit is at least 70% identical to SEQ ID NO:14.
6. The fusion protein of claim 5, wherein the peroxisome targeting protein subunit is at least 80% identical to SEQ ID NO:14.
7. The fusion protein of claim 6, wherein the peroxisome targeting protein subunit is at least 90% identical to SEQ ID NO:14.
8. The fusion protein of claim 7, wherein the peroxisome targeting protein subunit is SEQ ID NO:14.
9. The fusion protein of claim 1, wherein the polyhydroxyalkanoate synthase protein subunit is a *Pseudomonas* subunit.

10. The fusion protein of claim 9, wherein the *Pseudomonas* subunit is a *Pseudomonas aeruginosa* subunit.
11. The fusion protein of claim 10, wherein the polyhydroxyalkanoate synthase protein subunit is a PHAC1 subunit.
12. The fusion protein of claim 11, wherein the polyhydroxyalkanoate synthase protein subunit is at least 70% identical to SEQ ID NO:2.
13. The fusion protein of claim 12, wherein the polyhydroxyalkanoate synthase protein subunit is at least 80% identical to SEQ ID NO:2.
14. The fusion protein of claim 13, wherein the polyhydroxyalkanoate synthase protein subunit is at least 90% identical to SEQ ID NO:2.
15. The fusion protein of claim 14, wherein the polyhydroxyalkanoate synthase protein subunit is SEQ ID NO:2.
16. The fusion protein of claim 10, wherein the polyhydroxyalkanoate synthase protein subunit is a PHAC2 subunit.
17. The fusion protein of claim 16, wherein the polyhydroxyalkanoate synthase protein subunit is at least 70% identical to SEQ ID NO:4.
18. The fusion protein of claim 17, wherein the polyhydroxyalkanoate synthase protein subunit is at least 80% identical to SEQ ID NO:4.
19. The fusion protein of claim 18, wherein the polyhydroxyalkanoate synthase protein subunit is at least 90% identical to SEQ ID NO:4.
20. The fusion protein of claim 19, wherein the polyhydroxyalkanoate synthase protein subunit is SEQ ID NO:4.

21. The fusion protein of claim 1, wherein the polyhydroxyalkanoate synthase protein subunit is at least 70% identical to SEQ ID NO:18 or SEQ ID NO:20.
22. The fusion protein of claim 21, wherein the polyhydroxyalkanoate synthase protein subunit is at least 80% identical to SEQ ID NO:18 or SEQ ID NO:20.
23. The fusion protein of claim 22, wherein the polyhydroxyalkanoate synthase protein subunit is at least 90% identical to SEQ ID NO:18 or SEQ ID NO:20.
24. The fusion protein of claim 23, comprising SEQ ID NO:18 or SEQ ID NO:20.
25. A nucleic acid segment encoding a non-naturally occurring fusion protein, the nucleic acid segment comprising:  
a nucleic acid sequence encoding a peroxisome targeting protein subunit; and  
a nucleic acid sequence encoding a polyhydroxyalkanoate synthase protein subunit.
26. The nucleic acid segment of claim 25, wherein the nucleic acid sequence encoding a peroxisome targeting protein subunit comprises at least a 6 contiguous nucleic acid sequence from SEQ ID NO:13.
27. The nucleic acid segment of claim 25, wherein the nucleic acid sequence encoding a peroxisome targeting protein subunit is at least 70% identical to SEQ ID NO:13.
28. The nucleic acid segment of claim 27, wherein the nucleic acid sequence encoding a peroxisome targeting protein subunit is at least 80% identical to SEQ ID NO:13.
29. The nucleic acid segment of claim 28, wherein the nucleic acid sequence encoding a peroxisome targeting protein subunit is at least 90% identical to SEQ ID NO:13.
30. The nucleic acid segment of claim 29, wherein the nucleic acid sequence encoding a peroxisome targeting protein subunit is SEQ ID NO:13.

31. The nucleic acid segment of claim 25, wherein the nucleic acid sequence encoding a peroxisome targeting protein subunit hybridizes to SEQ ID NO:13.

32. The nucleic acid segment of claim 25, wherein the nucleic acid sequence encoding a polyhydroxyalkanoate synthase protein subunit comprises at least a 6 contiguous nucleic acid sequence from:

SEQ ID NO:1;  
SEQ ID NO:3;  
SEQ ID NO:15; or  
SEQ ID NO:16.

33. The nucleic acid segment of claim 25, wherein the nucleic acid sequence encoding a polyhydroxyalkanoate synthase protein subunit is at least 70% identical to:

SEQ ID NO:1;  
SEQ ID NO:3;  
SEQ ID NO:15; or  
SEQ ID NO:16.

34. The nucleic acid segment of claim 33, wherein the nucleic acid sequence encoding a polyhydroxyalkanoate synthase protein subunit is at least 80% identical to:

SEQ ID NO:1;  
SEQ ID NO:3;  
SEQ ID NO:15; or  
SEQ ID NO:16.

35. The nucleic acid segment of claim 34, wherein the nucleic acid sequence encoding a polyhydroxyalkanoate synthase protein subunit is at least 90% identical to:

SEQ ID NO:1;  
SEQ ID NO:3;  
SEQ ID NO:15; or  
SEQ ID NO:16.

36. The nucleic acid segment of claim 35, wherein the nucleic acid sequence encoding a polyhydroxyalkanoate synthase protein subunit is:

SEQ ID NO:1;  
SEQ ID NO:3;  
SEQ ID NO:15; or  
SEQ ID NO:16.

37. The nucleic acid segment of claim 36, wherein the nucleic acid sequence encoding a polyhydroxyalkanoate synthase protein subunit is:

SEQ ID NO:15; or  
SEQ ID NO:16.

38. The nucleic acid segment of claim 25, wherein the nucleic acid sequence encoding a polyhydroxyalkanoate synthase protein subunit hybridizes to:

SEQ ID NO:1;  
SEQ ID NO:3;  
SEQ ID NO:15; or  
SEQ ID NO:16.

39. The nucleic acid segment of claim 25, wherein the peroxisome targeting protein subunit is PTS2.

40. The nucleic acid segment of claim 25, wherein the peroxisome targeting protein subunit comprises a tripeptide, the tripeptide having:

a first amino acid in the N-terminus to C-terminus direction being S, A, or P;  
a second amino acid in the N-terminus to C-terminus direction being K, R, S, or H;  
and  
a third amino acid in the N-terminus to C-terminus direction being L, M, I, or F.

41. The nucleic acid segment of claim 40, wherein the peroxisome targeting subunit comprises ARM, SRM, SKL, ARL, SRL, PSL, or PRM.

42. The nucleic acid segment of claim 25, wherein the nucleic acid sequence encoding a polyhydroxyalkanoate synthase protein subunit encodes at least a 5 contiguous amino acid sequence from:

SEQ ID NO:2; or  
SEQ ID NO:4.

5

43. The nucleic acid segment of claim 25, wherein the nucleic acid sequence encoding a polyhydroxyalkanoate synthase protein subunit encodes an amino acid sequence at least about 70% identical to:

SEQ ID NO:2; or  
SEQ ID NO:4.

10

44. The nucleic acid segment of claim 43, wherein the nucleic acid sequence encoding a polyhydroxyalkanoate synthase protein subunit encodes an amino acid sequence at least about 80% identical to:

SEQ ID NO:2; or  
SEQ ID NO:4.

15

45. The nucleic acid segment of claim 44, wherein the nucleic acid sequence encoding a polyhydroxyalkanoate synthase protein subunit encodes an amino acid sequence at least about 90% identical to:

SEQ ID NO:2; or  
SEQ ID NO:4.

20

46. The nucleic acid segment of claim 45, wherein the nucleic acid sequence encoding a polyhydroxyalkanoate synthase protein subunit encodes:

SEQ ID NO:2; or  
SEQ ID NO:4.

25

47. A recombinant vector comprising in the 5' to 3' direction:

30

a) a promoter that directs transcription of a structural nucleic acid sequence encoding a non-naturally occurring fusion protein, wherein the fusion protein comprises:

i) a peroxisome targeting protein subunit; and  
ii) a polyhydroxyalkanoate synthase protein subunit.

5

b) a structural nucleic acid sequence encoding a non-naturally occurring fusion protein, wherein the fusion protein comprises:

i) a peroxisome targeting protein subunit; and  
ii) a polyhydroxyalkanoate synthase protein subunit; and

c) a 3' transcription terminator.

10

48. The recombinant vector of claim 47, further comprising a 3' polyadenylation signal sequence that directs the addition of polyadenylate nucleotides to the 3' end of RNA transcribed from the structural nucleic acid coding sequence.

15

49. The recombinant vector of claim 47, further comprising a selectable marker.

50. The recombinant vector of claim 49, wherein the selectable marker is a kanamycin resistance marker, a hygromycin resistance marker, or a herbicide resistance marker.

20

51. The recombinant vector of claim 47, wherein the promoter is constitutive.

52. The recombinant vector of claim 51, wherein the promoter is CaMV35S, enhanced CaMV35S, FMV, mas, nos, or ocs.

53. The recombinant vector of claim 47, wherein the promoter is inducible.

25

54. The recombinant vector of claim 53, wherein the promoter is tac, salicylic acid induced, polyacrylic acid induced, safener induced, heat shock promoter, nitrate induced, hormone induced, or light induced.

55. The recombinant vector of claim 47, wherein the promoter is tissue specific.

30

56. The recombinant vector of claim 55, wherein the promoter is the  $\beta$ -conglycinin 7S promoter, napin promoter, phaseolin promoter, zein promoter, soybean trypsin inhibitor promoter, ACP promoter, stearyl-ACP desaturase promoter, or oleosin promoter.
57. The recombinant vector of claim 47, wherein the nucleic acid sequence encoding a peroxisome targeting protein subunit comprises at least a 6 contiguous nucleic acid sequence from SEQ ID NO:13.
58. The recombinant vector of claim 47, wherein the nucleic acid sequence encoding a peroxisome targeting protein subunit is at least 70% identical to SEQ ID NO:13.
59. The recombinant vector of claim 58, wherein the nucleic acid sequence encoding a peroxisome targeting protein subunit is at least 80% identical to SEQ ID NO:13.
60. The recombinant vector of claim 59, wherein the nucleic acid sequence encoding a peroxisome targeting protein subunit is at least 90% identical to SEQ ID NO:13.
61. The recombinant vector of claim 60, wherein the nucleic acid sequence encoding a peroxisome targeting protein subunit is SEQ ID NO:13.
62. The recombinant vector of claim 47, wherein the nucleic acid sequence encoding a peroxisome targeting protein subunit hybridizes to SEQ ID NO:13.
63. The recombinant vector of claim 47, wherein the peroxisome targeting protein subunit is PTS2.
64. The recombinant vector of claim 47, wherein the peroxisome targeting protein subunit comprises a tripeptide, the tripeptide having:  
a first amino acid in the N-terminus to C-terminus direction being S, A, or P;  
a second amino acid in the N-terminus to C-terminus direction being K, R, S, or H;  
and  
a third amino acid in the N-terminus to C-terminus direction being L, M, I, or F.
65. The recombinant vector of claim 64, wherein the peroxisome targeting subunit comprises ARM, SRM, SKL, ARL, SRL, PSI, or PRM.
66. The recombinant vector of claim 47, wherein the polyhydroxyalkanoate synthase protein subunit is a *Pseudomonas* subunit.
67. The recombinant vector of claim 66, wherein the *Pseudomonas* subunit is a *Pseudomonas aeruginosa* subunit.
68. The recombinant vector of claim 47, wherein the nucleic acid sequence encoding a polyhydroxyalkanoate synthase protein subunit comprises at least a 6 contiguous nucleic acid sequence from:  
SEQ ID NO:1;  
SEQ ID NO:3;  
SEQ ID NO:15; or  
SEQ ID NO:16.
69. The recombinant vector of claim 47, wherein the nucleic acid sequence encoding a polyhydroxyalkanoate synthase protein subunit is at least 70% identical to:  
SEQ ID NO:1;  
SEQ ID NO:3;  
SEQ ID NO:15; or  
SEQ ID NO:16.
70. The recombinant vector of claim 69, wherein the nucleic acid sequence encoding a polyhydroxyalkanoate synthase protein subunit is at least 80% identical to:  
SEQ ID NO:1;  
SEQ ID NO:3;  
SEQ ID NO:15; or  
SEQ ID NO:16.

71. The recombinant vector of claim 70, wherein the nucleic acid sequence encoding a polyhydroxyalkanoate synthase protein subunit is at least 90% identical to:

SEQ ID NO:1;  
SEQ ID NO:3;  
SEQ ID NO:15; or  
SEQ ID NO:16.

5

72. The recombinant vector of claim 71, wherein the nucleic acid sequence encoding a polyhydroxyalkanoate synthase protein subunit is:

SEQ ID NO:1;  
SEQ ID NO:3;  
SEQ ID NO:15; or  
SEQ ID NO:16.

10

73. The recombinant vector of claim 72, wherein the nucleic acid sequence encoding a polyhydroxyalkanoate synthase protein subunit is:  
SEQ ID NO:15; or  
SEQ ID NO:16.

15

74. The recombinant vector of claim 47, wherein the nucleic acid sequence encoding a polyhydroxyalkanoate synthase protein subunit hybridizes to:  
SEQ ID NO:1;  
SEQ ID NO:3;  
SEQ ID NO:15; or  
SEQ ID NO:16.

20

25

75. The recombinant vector of claim 47, wherein the nucleic acid sequence encoding a polyhydroxyalkanoate synthase protein subunit encodes at least a 5 contiguous amino acid sequence from:  
SEQ ID NO:2; or  
SEQ ID NO:4.

30

76. The recombinant vector of claim 47, wherein the nucleic acid sequence encoding a polyhydroxyalkanoate synthase protein subunit encodes an amino acid sequence at least about 70% identical to:

SEQ ID NO:2; or  
SEQ ID NO:4.

5

77. The recombinant vector of claim 76, wherein the nucleic acid sequence encoding a polyhydroxyalkanoate synthase protein subunit encodes an amino acid sequence at least about 80% identical to:

SEQ ID NO:2; or  
SEQ ID NO:4.

10

78. The recombinant vector of claim 77, wherein the nucleic acid sequence encoding a polyhydroxyalkanoate synthase protein subunit encodes an amino acid sequence at least about 90% identical to:

SEQ ID NO:2; or  
SEQ ID NO:4.

15

79. The recombinant vector of claim 78, wherein the nucleic acid sequence encoding a polyhydroxyalkanoate synthase protein subunit encodes:

SEQ ID NO:2; or  
SEQ ID NO:4.

20

80. The recombinant vector of claim 47, wherein the structural nucleic acid sequence comprises:

SEQ ID NO:17; or  
SEQ ID NO:19.

81. The recombinant vector of claim 47, wherein the structural nucleic acid sequence encodes:

SEQ ID NO:18; or  
SEQ ID NO:20.

25

82. A recombinant host cell comprising a nucleic acid segment encoding a non-naturally occurring fusion protein, wherein the nucleic acid segment comprises:  
a nucleic acid sequence encoding a peroxisome targeting protein subunit; and  
a nucleic acid sequence encoding a polyhydroxyalkanoate synthase protein subunit.

83. The recombinant host cell of claim 82, wherein the recombinant host cell is a fungal cell.

84. The recombinant host cell of claim 83, wherein the fungal cell is a *Schizosaccharomyces pombe*, *Streptomyces rimofaciens*, *Fusarium*, *Aspergillus niger*, or *Saccharomyces cerevisiae* cell.

85. The recombinant host cell of claim 82, wherein the recombinant host cell is a plant cell.

86. The recombinant host cell of claim 85, wherein the plant cell is an alfalfa, banana, barley, bean, cabbage, canola/oilseed rape, carrot, castorbean, celery, clover, coconut, corn, cotton, cucumber, linseed, melon, olive, palm, parsnip, pea, peanut, pepper, potato, radish, rapeseed, rice, soybean, spinach, sunflower, tobacco, tomato, or wheat cell.

87. The recombinant host cell of claim 82, further comprising a nucleic acid segment encoding an acyl-ACP thioesterase.

88. The recombinant host cell of claim 82, further comprising a nucleic acid segment encoding a fatty acyl hydroxylase.

89. The recombinant host cell of claim 82, further comprising a nucleic acid segment encoding a yeast multifunctional protein (MFP).

90. The recombinant host cell of claim 82, further comprising a nucleic acid segment encoding a hydroxyacyl-CoA epimerase.

91. A genetically transformed plant cell comprising in the 5' to 3' direction:

a) a promoter to direct transcription of a structural nucleic acid sequence encoding a non-naturally occurring fusion protein, wherein the structural nucleic acid sequence comprises:

i) a nucleic acid sequence encoding a peroxisome targeting protein subunit; and

ii) a nucleic acid sequence encoding a polyhydroxyalkanoate synthase protein subunit;

b) a structural nucleic acid sequence encoding a non-naturally occurring fusion protein, wherein the structural nucleic acid sequence comprises:

i) a nucleic acid sequence encoding a peroxisome targeting protein subunit; and

ii) a nucleic acid sequence encoding a polyhydroxyalkanoate synthase protein subunit;

c) a 3' transcription terminator sequence; and

d) a 3' polyadenylation signal sequence that directs the addition of polyadenylate nucleotides to the 3' end of RNA transcribed from the structural nucleic acid coding sequence.

92. The genetically transformed plant cell of claim 91, wherein the plant cell is an alfalfa, banana, barley, bean, cabbage, canola/oilseed rape, carrot, castorbean, celery, clover, coconut, corn, cotton, cucumber, linseed, melon, olive, palm, parsnip, pea, peanut, pepper, potato, radish, rapeseed, rice, soybean, spinach, sunflower, tobacco, tomato, or wheat cell.

93. The genetically transformed plant cell of claim 91, further comprising a nucleic acid segment encoding an acyl-ACP thioesterase.



94. The genetically transformed plant cell of claim 91, further comprising a nucleic acid segment encoding a fatty acyl hydroxylase.

95. The genetically transformed plant cell of claim 91, further comprising a nucleic acid segment encoding a yeast multifunctional protein (MFP).

96. The genetically transformed plant cell of claim 91, further comprising a nucleic acid segment encoding a hydroxyacyl-CoA epimerase.

97. A genetically transformed plant comprising in the 5' to 3' direction:

a) a promoter to direct transcription of a structural nucleic acid sequence encoding a non-naturally occurring fusion protein, wherein the structural nucleic acid sequence comprises:

i) a nucleic acid sequence encoding a peroxisome targeting protein subunit; and

ii) a nucleic acid sequence encoding a polyhydroxyalkanoate synthase protein subunit;

b) a structural nucleic acid sequence encoding a non-naturally occurring fusion protein, wherein the structural nucleic acid sequence comprises:

i) a nucleic acid sequence encoding a peroxisome targeting protein subunit; and

ii) a nucleic acid sequence encoding a polyhydroxyalkanoate synthase protein subunit;

c) a 3' transcription terminator sequence; and

d) a 3' polyadenylation signal sequence that directs the addition of polyadenylate nucleotides to the 3' end of RNA transcribed from the structural nucleic acid coding sequence.

98. The genetically transformed plant of claim 97, wherein the plant is an alfalfa, banana, barley, bean, cabbage, canola/oilseed rape, carrot, castorbean, celery, clover, coconut, corn, cotton, cucumber, linseed, melon, olive, palm, parsnip, pea, peanut,

pepper, potato, radish, rapeseed, rice, soybean, spinach, sunflower, tobacco, tomato, or wheat plant.

99. The genetically transformed plant of claim 97, wherein the promoter is constitutive.

100. The genetically transformed plant of claim 99, wherein the promoter is CaMV35S, enhanced CaMV35S, FMV, mas, nos, or ocs.

101. The genetically transformed plant of claim 97, wherein the promoter is inducible.

102. The genetically transformed plant of claim 101, wherein the promoter is tac, salicylic acid induced, polyacrylic acid induced, safener induced, heat shock promoter, nitrate induced, hormone induced, or light induced.

103. The genetically transformed plant of claim 97, wherein the promoter is tissue specific.

104. The genetically transformed plant of claim 103, wherein the promoter is the  $\beta$ -conglycinin 7S promoter, napin promoter, phaseolin promoter, zein promoter, soybean trypsin inhibitor promoter, ACP promoter, stearoyl-ACP desaturase promoter, or oleosin promoter.

105. The genetically transformed plant of claim 97, further comprising a nucleic acid segment encoding an acyl-ACP thioesterase.

106. The genetically transformed plant of claim 97, further comprising a nucleic acid segment encoding a fatty acyl hydroxylase.

107. The genetically transformed plant of claim 97, further comprising a nucleic acid segment encoding a yeast multifunctional protein (MFP).

108. The genetically transformed plant of claim 97, further comprising a nucleic acid segment encoding a hydroxyacyl-CoA epimerase.

109. A method of preparing host cells useful to produce a non-naturally occurring fusion protein comprising the steps of:

- a) selecting a host cell
- b) transforming the selected host cell with a recombinant vector having a structural nucleic acid sequence encoding a non-naturally occurring fusion protein, wherein the structural nucleic acid sequence comprises:
  - i) a nucleic acid sequence encoding a peroxisome targeting protein subunit; and
  - ii) a nucleic acid sequence encoding a polyhydroxyalkanoate synthase protein subunit; and
- c) obtaining transformed host cells.

110. The method of claim 109, wherein the vector further comprises a selectable marker.

111. The method of claim 110, wherein the selectable marker is a kanamycin resistance marker, a hygromycin resistance marker, or a herbicide resistance marker.

112. The method of claim 109, wherein the host cell is a fungal cell.

113. The method of claim 112, wherein the fungal cell is a *Schizosaccharomyces pombe*, *Streptomyces rimofaciens*, *Fusarium*, *Aspergillus niger*, or *Saccharomyces cerevisiae* cell.

114. The method of claim 109, wherein the host cell is a plant cell.

115. The method of claim 114, wherein the plant cell is an alfalfa, banana, barley, bean, cabbage, canola/oilseed rape, carrot, castorbean, celery, clover, coconut, corn, cotton, cucumber, linseed, melon, olive, palm, parsnip, pea, peanut, pepper, potato,

potato, radish, rapeseed, rice, soybean, spinach, sunflower, tobacco, tomato, or wheat cell.

116. A method of preparing a transformed plant useful to produce a non-naturally occurring fusion protein comprising the steps of:

- a) selecting a host plant cell
- b) transforming the selected host plant cell with a recombinant vector having a structural nucleic acid sequence encoding a non-naturally occurring fusion protein, wherein the structural nucleic acid sequence comprises:
  - i) a nucleic acid sequence encoding a peroxisome targeting protein subunit; and
  - ii) a nucleic acid sequence encoding a polyhydroxyalkanoate synthase protein subunit;
- c) obtaining transformed host plant cells; and
- d) regenerating the transformed host plant cells.

117. The method of claim 116, wherein the vector further comprises a selectable marker.

118. The method of claim 117, wherein the selectable marker is a kanamycin resistance marker, a hygromycin resistance marker, or a herbicide resistance marker.

119. The method of claim 116, wherein the host plant cell is an alfalfa, banana, barley, bean, cabbage, canola/oilseed rape, carrot, castorbean, celery, clover, coconut, corn, cotton, cucumber, linseed, melon, olive, palm, parsnip, pea, peanut, pepper, potato, potato, radish, rapeseed, rice, soybean, spinach, sunflower, tobacco, tomato, or wheat cell.

120. The plant produced by the method of claim 116.

121. A method for the preparation of a polyhydroxyalkanoate, comprising the steps of:

- a) obtaining a cell capable of producing a non-naturally occurring fusion protein, wherein the fusion protein comprises:
  - i) a peroxisome targeting protein subunit; and
  - ii) a polyhydroxyalkanoate synthase protein subunit;
- b) establishing a culture of the cell; and
- c) culturing the cell under conditions suitable for the production of the polyhydroxyalkanoate.

122. The method of claim 121, wherein the culture contains natural fatty acids, non-natural fatty acids, or mixtures thereof.

123. The method of claim 121, wherein the cell is a fungal cell.

124. The method of claim 123, wherein the fungal cell is a *Schizosaccharomyces pombe*, *Streptomyces rimofaciens*, *Fusarium*, *Aspergillus niger*, or *Saccharomyces cerevisiae* cell.

125. The method of claim 121, wherein the cell is a plant cell.

126. The method of claim 125, wherein the cell is an alfalfa, banana, barley, bean, cabbage, canola/oilseed rape, carrot, castorbean, celery, clover, coconut, corn, cotton, cucumber, linseed, melon, olive, palm, parsnip, pea, peanut, pepper, potato, potato, radish, rapeseed, rice, soybean, spinach, sunflower, tobacco, tomato, or wheat cell.

127. The method of claim 121, wherein the polyhydroxyalkanoate comprises 3-hydroxyhexanoic acid (H:6), 3-hydroxyoctanoic acid (H:8), 3-hydroxydecanoic acid (H:10), 3-hydroxydodecanoic acid (H:12), 3-hydroxytetradecanoic acid (H:14), 3-hydroxyhexadecanoic acid (H:16), 3-hydroxyheptanoic acid (H:7), 3-hydroxynonanoic acid (H:9), 3-hydroxyundecanoic acid (H:11), 3-

hydroxytridecanoic acid (H:13), 3-hydroxyhexadecatrienoic acid (H:16:3), 3-hydroxyhexadecadienoic acid (H:16:2), 3-hydroxyhexadecenoic acid (H:16:1), 3-hydroxytetradecatrienoic acid (H:14:3), 3-hydroxytetradecadienoic acid (H:14:2), 3-hydroxytetradecenoic acid (H:14:1), 3-hydroxydodecadienoic acid (H:12:2), 3-hydroxydodecenoic acid (H:12:1), 3-hydroxyoctenoic acid (H:8:1), 4-hydroxydecanoic acid, 8-methyl-3-hydroxynonanoic acid, or 6-methyl-3-hydroxyheptanoic acid monomers.

128. A method for the preparation of a polyhydroxyalkanoate, comprising the steps of:

- a) obtaining a plant capable of producing a non-naturally occurring fusion protein, wherein the fusion protein comprises:
  - i) a peroxisome targeting protein subunit; and
  - ii) a polyhydroxyalkanoate synthase protein subunit; and
- b) growing the plant under conditions suitable for the production of the polyhydroxyalkanoate.

129. The method of claim 128, further comprising supplementing the plant with natural fatty acids, non-natural fatty acids, or mixtures thereof.

130. The method of claim 128, wherein the plant is an alfalfa, banana, barley, bean, cabbage, canola/oilseed rape, carrot, castorbean, celery, clover, coconut, corn, cotton, cucumber, linseed, melon, olive, palm, parsnip, pea, peanut, pepper, potato, potato, radish, rapeseed, rice, soybean, spinach, sunflower, tobacco, tomato, or wheat plant.

131. The method of claim 128, wherein the polyhydroxyalkanoate comprises 3-hydroxyhexanoic acid (H:6), 3-hydroxyoctanoic acid (H:8), 3-hydroxydecanoic acid (H:10), 3-hydroxydodecanoic acid (H:12), 3-hydroxytetradecanoic acid (H:14), 3-hydroxyhexadecanoic acid (H:16), 3-hydroxyheptanoic acid (H:7), 3-hydroxynonanoic acid (H:9), 3-hydroxyundecanoic acid (H:11), 3-hydroxytridecanoic acid (H:13), 3-hydroxyhexadecatrienoic acid (H:16:3), 3-

hydroxyhexadecadienoic acid (H16:2), 3-hydroxyhexadecenoic acid (H16:1), 3-hydroxytetradecatrienoic acid (H14:3), 3-hydroxytetradecadienoic acid (H14:2), 3-hydroxytetradecenoic acid (H14:1), 3-hydroxydodecadienoic acid (H12:2), 3-hydroxydodecenoic acid (H12:1), 3-hydroxyoctenoic acid (H8:1), 4-hydroxydecanoic acid, 8-methyl-3-hydroxynonanoic acid, or 6-methyl-3-hydroxyheptanoic acid monomers.

132. A plant containing a polyhydroxyalkanoate, wherein the polyhydroxyalkanoate comprises 3-hydroxyhexanoic acid (H:6), 3-hydroxyoctanoic acid (H:8), 3-hydroxydecanoic acid (H:10), 3-hydroxydodecanoic acid (H:12), 3-hydroxytetradecanoic acid (H:14), 3-hydroxyhexadecanoic acid (H:16), 3-hydroxyheptanoic acid (H:7), 3-hydroxynonanoic acid (H:9), 3-hydroxyundecanoic acid (H:11), 3-hydroxytridecanoic acid (H:13), 3-hydroxyhexadecatrienoic acid (H16:3), 3-hydroxyhexadecadienoic acid (H16:2), 3-hydroxyhexadecenoic acid (H16:1), 3-hydroxytetradecatrienoic acid (H14:3), 3-hydroxytetradecadienoic acid (H14:2), 3-hydroxytetradecenoic acid (H14:1), 3-hydroxydodecadienoic acid (H12:2), 3-hydroxydodecenoic acid (H12:1), 3-hydroxyoctenoic acid (H8:1), 4-hydroxydecanoic acid, 8-methyl-3-hydroxynonanoic acid, or 6-methyl-3-hydroxyheptanoic acid monomers.

133. A polyhydroxyalkanoate comprising 3-hydroxyhexadecatrienoic acid (H16:3), 3-hydroxyhexadecadienoic acid (H16:2), 3-hydroxytetradecatrienoic acid (H14:3), or 3-hydroxydodecadienoic acid (H12:2) monomers.

1/2

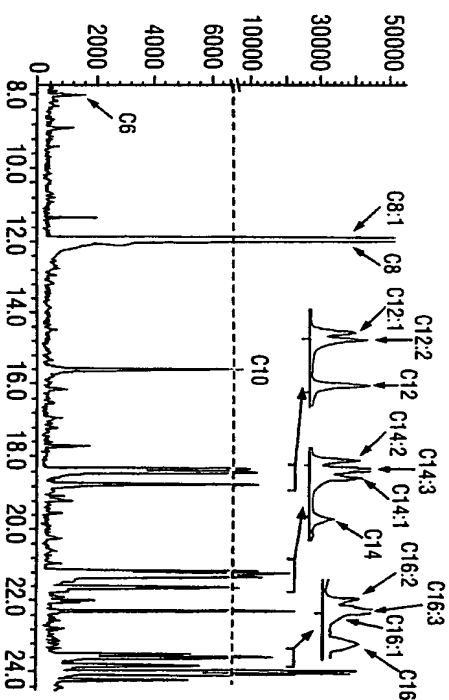


FIG. 1A

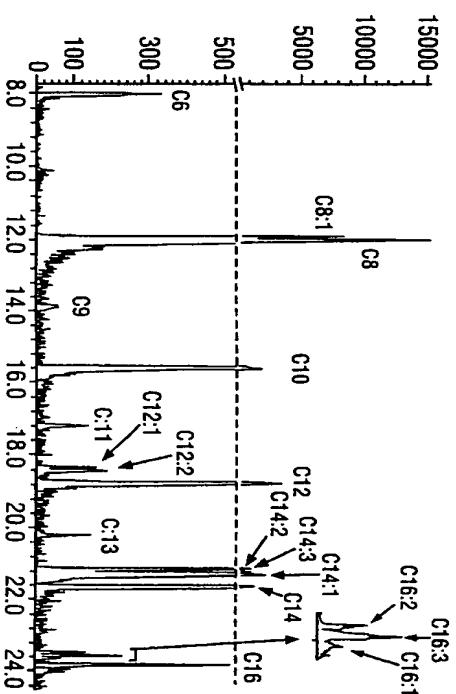


FIG. 1B

2/2

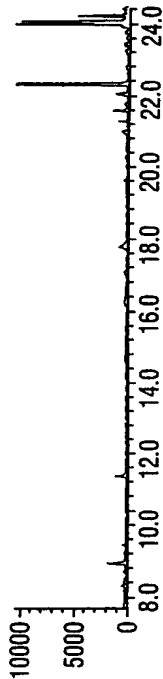


FIG. 1C

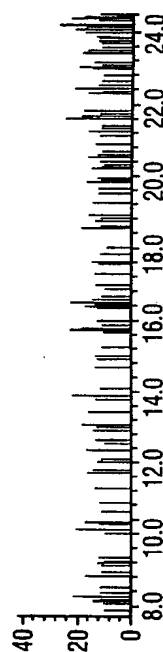


FIG. 1D

## INTERNATIONAL SEARCH REPORT

International Application No. PCT/US 98/00083	
A. CLASSIFICATION OF SUBJECT MATTER IPC 6 C12N15/82 C12N15/52 C12N15/62 C12N15/81 C12P7/62 A01H5/00 C08G63/06	
According to International Patent Classification (IPC) or to both national classification and IPC	
B. FIELDS SEARCHED Minimum documentation searched (classification system followed by classification symbols) IPC 6 C12P C12N A01H C08G	
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched	
Electronic data base consulted during the international search (name of data base and, where practical, search terms used)	
C. DOCUMENTS CONSIDERED TO BE RELEVANT	
Category <sup>1</sup>	Citation of document, with indication, where appropriate, of the relevant passages Relevant to claim No.
X	EP 0 526 850 A (KANEGAFUCHI CHEMICAL IND) 10 February 1993 see the whole document 133
X	WO 95 33064 A (PROCTER & GAMBLE) 7 December 1995 see examples 1-4 132
X	WO 95 33065 A (PROCTER & GAMBLE) 7 December 1995 see examples 1-5 132
X	WO 97 07230 A (PROCTER & GAMBLE) 27 February 1997 see examples 1-3 132
	--- -/-
<input checked="" type="checkbox"/> Further documents are listed in the continuation of box C. <input checked="" type="checkbox"/> Patent family members are listed in annex.	
<sup>1</sup> Special categories of cited documents: "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier document but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another claim or other special reason (as specified) "O" document relating to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed	
Date of the actual completion of the international search 17 September 1998	
Date of mailing of the international search report 29/09/1998	
Name and mailing address of the ISA European Patent Office, P.B. 5618 Patentlaan 2 NL-2580 RT The Hague Tel. (+31-70) 340-2040, Tx. 31 651 gpo nl. Fax: (+31-70) 340-2016	
Authorized officer Maddox, A	

International Application No.  
PCT/US 98/00083

Category		Category	Category
Citation of document, with indication where appropriate, of the relevant passages		Relevant to claim No.	
X	WO 97 07229 A (PROCTER & GAMBLE) 27 February 1997 see example 4	132	
Y	WO 94 11519 A (ZENECA LTD.; FENTEM PHILLIP ANTHONY (GB)) 26 May 1994	1-82, 85, 86, 91, 92, 97-104, 109-111, 114-122, 125-131	
Y	see page 5, line 17 - page 6, line 34 see page 8, line 21 - page 11, line 19	1-82, 85, 86, 91, 92, 97-104, 109-111, 114-122, 125-131	
Y	TIMM, A., ET AL.: "Cloning and molecular analysis of the poly(3-hydroxyalkanoic acid) gene locus of <i>Pseudomonas aeruginosa</i> PAO1" EUROPEAN JOURNAL OF BIOCHEMISTRY, vol. 209, no. 1, October 1992, pages 15-30, XP002077729 see figure 2	3, 4, 64, 65	
Y	HAYASHI, M., ET AL.: "Changes in the targeting efficiencies of proteins to plant microbodies caused by amino acid substitutions in the carboxy-terminal tripeptide" PLANT CELL PHYSIOL., vol. 38, no. 6, 1997, pages 759-768, XP002077730 see the whole document	5-8, 26-31, 57-62, 80, 81	
Y	OLSEN, L.J., ET AL.: "Targeting of glyoxysomal proteins to peroxisomes in leaves and roots of a higher plant" THE PLANT CELL, vol. 5, no. 8, August 1993, pages 941-952, XP002077731 see the whole document	2, 39, 63	
Y	KATO, A., ET AL.: "Targeting and processing of a chimeric protein with the N-terminal presence of the precursor to glyoxysomal citrate synthase" THE PLANT CELL, vol. 8, no. 9, September 1996, pages 1601-1611, XP002077732 see the whole document		

International Application No.  
PCT/US 98/00083

Category	C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT	Relevant to claim No.
<p>Children of document, with indication where appropriate, of the relevant passages</p>	<p>3, 4</p>	<p>5-8, 26-31, 57-62, 80, 81</p>
<p>HAYASHI, M., ET AL.: "Transport of chimeric proteins that contain a carboxy-terminal targeting signal into plant microbodies." THE PLANT JOURNAL, vol. 10, no. 2, 1996, pages 225-234. XP002077733 see the whole document</p>	<p>5-8, 26-31, 57-62, 80, 81</p>	<p>5-8, 26-31, 57-62, 80, 81</p>
<p>OLESEN, C., ET AL.: "Brassica napus mRNA for glyoxysomal isocitrate lyase" EMBL SEQUENCE ACCESSION NO. Y13356, 29 May 1997. XP002077734 see the whole document</p>	<p>12, 32-34, 38</p>	<p>12, 32-34, 38</p>
<p>WO 91 00917 A (MASSACHUSETTS INST TECHNOLOG) 24 January 1991 see page 36, line 15 - page 38, line 6; figure 6 see page 45, line 23 - page 47, line 30</p>	<p>82, 84, 109, 112, 113</p>	<p>82, 84, 109, 112, 113</p>
<p>BIOLOGICAL ABSTRACTS, vol. 102, Philadelphia, PA, US; abstract no. 4612, LEAF T A, ET AL.: "Saccharomyces cerevisiae expressing bacterial polyhydroxybutyrate synthase produces poly-3-hydroxybutyrate." XP002077735 see abstract &amp; MICROBIOLOGY (READING) 142 (5), 1996, 1169-1180.</p>	<p>82, 84, 109, 112, 113</p>	<p>82, 84, 109, 112, 113</p>
<p>WO 94 24289 A (EUROLYSINE ;NICAUD JEAN MARC (FR); LABAT NATHALIE (FR); ITO HISAO) 27 October 1994 see the whole document</p>	<p>87-90</p>	<p>87-90</p>
<p>WO 97 22711 A (UNIV MINNESOTA ;SHERMAN DAVID H (US); WILLIAMS MARK D (US); XUE Y0) 26 June 1997 see the whole document</p>	<p>1-133</p>	<p>1-133</p>
<p>LEIJ VAN DER F R ET AL.: "STRATEGIES FOR THE SUSTAINABLE PRODUCTION OF NEW BIODEGRADABLE POLYESTERS IN PLANTS: A REVIEW" CANADIAN JOURNAL OF MICROBIOLOGY, vol. 41, no. SUPPL. 01, 1995, pages 222-238. XP002045278 see page 232 - page 235</p>	<p>1-133</p>	<p>1-133</p>

# INTERNATIONAL SEARCH REPORT

INTERNATIONAL SEARCH REPORT		International Application No. PCT/US 98/00083
C(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		Relevant to claim No.
Category	Citation of document, with indication, where appropriate, of the relevant passages	
A	POIRIER Y ET AL: "PRODUCTION OF POLYHYDROXYALKANOATES, A FAMILY OF BIODEGRADABLE PLASTICS AND ELASTOMERS, IN BACTERIA AND PLANTS" BIOTECHNOLOGY, vol. 13, February 1995, pages 142-150, XP002045222 see the whole document	1-133
A	EGGINK G ET AL: "THE ROLE OF FATTY ACID BIOSYNTHESIS AND DEGRADATION IN THE SUPPLY OF SUBSTRATES FOR POLY(3-HYDROXYALKANOATE) FORMATION IN PSEUDOMONASPUTIDA" FEMS MICROBIOLOGY REVIEWS, vol. 103, no. 2/04, December 1992, pages 159-163, XP002045279 see page 161, right-hand column - page 162	1-133
A	EP 0 274 151 A (UNIV GRONINGEN) 13 July 1988 see the whole document	1-133
A	WO 92 19747 A (ICI PLC) 12 November 1992 see the whole document	1-133
A	WO 95 05472 A (UNIV MICHIGAN) 23 February 1995 see the whole document	1-133
A	WO 93 02187 A (UNIV MICHIGAN) 4 February 1993 see the whole document	1-133
A	WO 92 18553 A (KOHAP LTD) 29 October 1992 see the whole document	132, 133

# INTERNATIONAL SEARCH REPORT

INTERNATIONAL SEARCH REPORT		International Application No. PCT/US 98/00083
Information on patent family members		
Patent document cited in search report	Publication date	Patent family member(s)
EP 0526850 A	10-02-1993	JP 5030980 A 09-02-1993 US 5290910 A 01-03-1994
WO 9533064 A	07-12-1995	AU 2649795 A 21-12-1995 BR 9507814 A 12-08-1997 CA 2191570 A 07-12-1995 CN 1152944 A 25-06-1997 EP 0763125 A 19-03-1997 FI 964786 A 29-01-1997 JP 10501133 T 03-02-1998
WO 9533065 A	07-12-1995	AU 2650095 A 21-12-1995 BR 9507830 A 16-09-1997 CA 2191713 A 07-12-1995 CN 1152943 A 25-06-1997 EP 0763126 A 19-03-1997 FI 964785 A 29-01-1997 JP 10501134 T 03-02-1998
WO 9707230 A	27-02-1997	AU 6847596 A 12-03-1997 NO 980702 A 21-04-1998
WO 9707229 A	27-02-1997	AU 6775896 A 12-03-1997 EP 0846184 A 10-06-1998 NO 980701 A 21-04-1998
WO 9411519 A	26-05-1994	AU 5427194 A 08-06-1994
WO 9100917 A	24-01-1991	CA 2062816 A 11-01-1991 EP 0482077 A 29-04-1992 JP 5500751 T 18-02-1993 US 5480794 A 02-01-1996 US 5534432 A 09-07-1996 US 5663063 A 02-09-1997 US 5245023 A 14-09-1993 US 5250430 A 05-10-1993
WO 9424289 A	27-10-1994	FR 2704237 A 28-10-1994 AU 6681094 A 08-11-1994

## INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No  
PCT/US 98/00083

Patent document cited in search report	Publication date	Patent family member(s)	Publication date	
WO 9722711	A	26-06-1997	NONE	
EP 0274151	A	13-07-1988	NL 8603073 A CA 1335370 A DE 3789377 D DE 3789377 T JP 2642937 B JP 63226291 A US 5135859 A US 5334698 A	01-07-1988 25-04-1995 21-04-1994 28-07-1994 20-08-1997 20-09-1988 04-08-1992 02-08-1994
WO 9219747	A	12-11-1992	AU 655816 B AU 1579792 A CA 2109221 A EP 0588988 A JP 6510422 T US 5502273 A	12-01-1995 21-12-1992 25-10-1992 06-04-1994 24-11-1994 26-03-1996
WO 9505472	A	23-02-1995	US 5610041 A AU 7633094 A EP 0719342 A JP 9501832 T	11-03-1997 14-03-1995 03-07-1996 25-02-1997
WO 9302187	A	04-02-1993	AU 671953 B AU 2323892 A BE 1006851 A BR 9206297 A CA 2113412 A CA 9400108 A CZ 9400108 A EP 0595896 A ES 2046142 A FI 940242 A FR 2679735 A HU 66825 A IT 1255348 B JP 6509233 T MX 9204245 A NL 9201288 A NO 940157 A NZ 243626 A	19-09-1996 23-02-1993 10-01-1995 17-10-1995 04-02-1993 15-12-1994 11-05-1994 16-01-1994 11-03-1994 05-02-1993 30-01-1995 31-10-1995 20-10-1994 01-02-1993 16-02-1993 18-03-1994 28-03-1995

From PCT/US 98/00083 (patent family member) (July 1993)

## INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No  
PCT/US 98/00083

Patent document cited in search report	Publication date	Patent family member(s)	Publication date	
WO 9302187	A	PL 170467 B PT 100705 A SK 6494 A US 5610041 A US 5650555 A	31-12-1996 29-10-1993 05-10-1994 11-03-1997 22-07-1997	
WO 9218553	A	29-10-1992	AT 154048 T AU 8612191 A AU 8612644 D DE 6912644 T DE 6912644 T EP 0582567 A JP 7500615 T US 5395919 A	15-06-1997 17-11-1992 10-07-1997 25-09-1997 16-02-1994 19-01-1995 07-03-1995

From PCT/US 98/00083 (patent family member) (July 1993)